Part III

H^∞ Theory

In this part of the book we discuss the general optimization problem OPT, which is the main mathematical tool on which our control design method is based. The discussion is restricted to scalar valued analytic functions. In control applications this corresponds to the SISO case.

In Chapter 9 the mathematical problem OPT is formulated, and its relevance to Control is treated. We introduce performance functions and sublevel sets of a performance function.

In Chapter 10 we give a test for determining if a given function is a solution to OPT. This one simple characterization of optima is the key to understanding OPT both conceptually and numerically. This test has a beautiful graphical interpretation that is also easy to implement on a computer.

Chapter 11 contains background facts on analytic functions and lists three fundamental properties of analytic functions, which every control engineer should know.

Chapter 12 gives the proof of the characterization given in Chapter 10 of solutions to OPT. The proof follows quickly from the three basic properties listed in Chapter 3.

Chapter 13 concerns computations of solutions to OPT. We begin with computer diagnostics, follow with a sample computer run, and then we discuss an algorithm for finding solutions.

Chapter 9

H^{∞}-Optimization and Control

9.1 The problem OPT

Let X be a subset of the complex plane C, and let $f : X \to C$ be a continuous function. The function f is said to be

- *bounded* (uniformly) if there exists $m > 0$ such that $|f(s)| < m$ for all $s \in X$. The symbol $\sup_{s \in X} |f(s)|$ denotes the smallest of such m.
- *real* (on the real axis) if it satisfies[1] $f(s) = \overline{f(\overline{s})}$.
- *analytic* if it can be expanded as a power series about any point s_0 in the interior of X and convergent in some (possibly small) disk centered at s_0.

We denote with A_{RHP} the set of all functions $f(s)$ that are real, bounded and continuous in the closed Right–Half–Plane (*R.H.P.*), and that are *analytic* on the open *R.H.P.* Therefore, $f(j\omega)$ is a bounded, continuous and real function of ω, for any $f(s) \in A_{RHP}$. The set A_{RHP} is a (real) vector space with the usual properties of addition

$$(f+g)(s) = f(s) + g(s)$$

and scalar multiplication

$$(\alpha f)(s) = \alpha f(s), \quad (\forall \alpha \in R)$$

---

[1] This definition makes sense only for sets X that are symmetric about the x-axis.

It is well known from the theory of analytic functions that an f continuous on a closed set and analytic on the interior of the set achieves the maximum modulus on the boundary of the set. As a consequence of this we have that any f in A_{RHP} satisfies

$$\sup_{\omega \in R} |f(j\omega)| = \sup_{s \in RHP} |f(s)|$$

A positive function of $\omega \in \mathbf{R}$ and $z \in C$ is called a *performance function.* We will use the notation $\Gamma(\omega, z)$ to denote a performance function. A performance function Γ that satisfies

$$\Gamma(\omega, z) = \Gamma(-\omega, \bar{z}) \tag{9.1}$$

is said to be *real symmetric.* Performance functions that arise from engineering applications are real symmetric, consequently all the performance functions considered in this book are real symmetric. So from now on, by the word "performance" what we really mean is "real symmetric performance".

The main optimization problem we consider[2] is,

OPT Given a performance function $\Gamma(\omega, z)$, find $\gamma^* \geq 0$ and f^* in A_{RHP} such that

$$\gamma^* = \inf_{f \in A_{RHP}} \sup_{\omega} \Gamma(\omega, f(j\omega)) = \sup_{\omega} \Gamma(\omega, f^*(j\omega)).$$

We will refer to the function f^* in the statement of OPT as a *solution* or *minimizer*, and refer to γ^* as the *optimal value.* Solutions f^* to OPT do not always exist in A_{RHP}, however for many performance functions they do (see Section 10.4).

9.2 The fundamental H^∞ problem of control

The OPT problem is central to the design of a system where specifications are given in the frequency domain and where stability is a key issue. Suppose that our objective is to design a linear, time invariant system in the frequency domain description, with stability a major consideration. Part of the system is given and we are forced to use (called the plant), and another part f we get to choose (designable part) (see Fig. 1). The stipulation that the designable part of the circuit be stable can be restated as f having no poles in the right half plane.

---

[2] As stated here, OPT corresponds to SISO system design. The case MIMO leads to OPT with admissible functions $\mathbf{f} = (f_1, \ldots, f_N)$, with each $f_\ell \in A_{RHP}$. See Chapter 15.

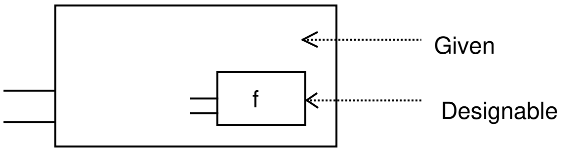

Figure 9.1: The given and designable parts of the system

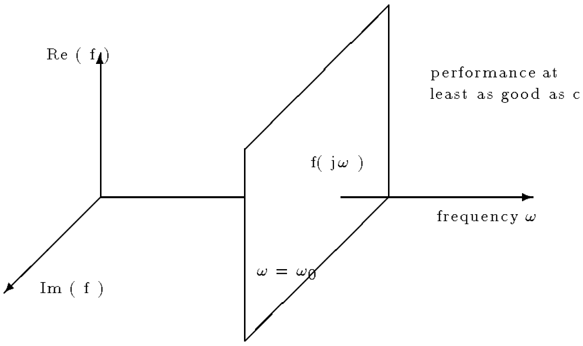

Figure 9.2: The function f makes performance at least as good as c at $\omega = \omega_0$.

Suppose also that to carry out the design (i.e., choose f) we use a given performance function Γ. In many cases the performance takes the form of a positive function of frequency ω and of the designable part f at that frequency.

Thus we write $\Gamma(\omega, f(j\omega))$ for the performance at frequency ω and for designable function f. The performance is a cost function: big is bad. The "worst case" is the frequency ω at which $\sup_{\omega \in R} \Gamma(\omega, f(j\omega))$ occurs. We minimize this over all $f \in A_{RHP}$. Thus we obtained exactly the OPT problem.

The theory of OPT is based in part on the analysis of the sublevel sets

$$\mathcal{S}_{\omega}(c) = \{z \in \mathbf{C}\colon \Gamma(\omega, z) < c\}$$

of Γ. In fact there is a simple picture to think of in connection with a design. For each desired level of performance c, one has target sets $\mathcal{S}_{\omega}(c)$ associated with each frequency ω. The objective is to find a function f with no poles in the R.H.P. such that each $f(j\omega)$ belongs to $\mathcal{S}_{\omega}(c)$. Any such f makes the performance of the overall system at least as good as c for all ω. See Fig. 9.2.

9.3 Change of coordinates from RHP to the unit disk

Because it is more convenient[3] we now change coordinates in the domains of functions, and use ζ in the unit disk D in the complex plane, and $e^{j\theta}$ in the unit circle instead of s in RHP, and $j\omega$ in the imaginary axis.

The change of coordinates can be accomplished with a linear fractional transformation:

$$\begin{aligned} s &= \tfrac{1+\zeta}{1-\zeta} \\ j\omega &= \tfrac{1+e^{j\theta}}{1-e^{j\theta}} \end{aligned} \tag{9.2}$$

so that functions $f(s)$ of s become functions of ζ:

$$\begin{aligned} f(s) &= f(\tfrac{1+\zeta}{1-\zeta}) &= F(\zeta) \\ f(j\omega) &= f(\tfrac{1+e^{j\theta}}{1-e^{j\theta}}) &= F(e^{j\theta}) \end{aligned} \tag{9.3}$$

The transformation (9.2) maps the open RHP one-to-one and onto the open unit disk D, and maps the $j\omega$ axis with $\omega = \infty$ one-to-one and onto the unit circle.

A fundamental property of this coordinate change is that functions remains analytic, bounded and real: $f(s)$ is analytic, bounded and real for $s \in RHP$ if and only if $F(\zeta)$ is analytic, bounded and real for $\zeta \in D$. The following table gives some pairs (s, ζ) that arise from this transformation.

s	∞	j	0	1	-1
ζ	1	j	-1	0	∞

Example. We perform the change of coordinates on the function $f(s) = (s-3)/(s+2)$. We have,

$$f(s) = \frac{s-3}{s+2} = \frac{\frac{1+\zeta}{1-\zeta} - 3}{\frac{1+\zeta}{1-\zeta} + 2} = \frac{\frac{-2+4\zeta}{1-\zeta}}{\frac{3-\zeta}{1-\zeta}} = \frac{-2+4\zeta}{3-\zeta} = F(\zeta) \tag{9.4}$$

---

[3] Mathematics literature treats functions on the disk rather the RHP. There are other reasons, e.g. our graphical test in Section 10 is easier on the disk, Fourier transforms and their discrete versions are available and easy to implement on the computer, etc.

Note that $f(s)$ has a pole in the L.H.P. and a zero in the R.H.P. Correspondingly, $F(\zeta)$ has a pole outside the unit disk and a zero inside the unit disk.

Instead of A_{RHP} we will use instead its "disk" counterpart that we denote by A. In other words, A is the set of all functions $f(\zeta)$ that are bounded, real and continuous on the closed unit disk, and analytic on the open unit disk. With the change of coordinates, the main optimization problem becomes

OPT Given a performance function $\Gamma(e^{j\theta}, z)$, find $\gamma^* \geq 0$ and f^* in A such that

$$\gamma^* = \inf_{f \in A} \sup_{\theta} \Gamma(e^{j\theta}, f(e^{j\theta})) = \sup_{\theta} \Gamma(e^{j\theta}, f^*(e^{j\theta})).$$

9.4 Performance functions and sublevel sets

With the coordinate change introduced in the previous subsection, the sublevel sets of $\Gamma(e^{j\theta}, z)$ now depend on θ:

$$S_\theta(c) = \{z : \Gamma(e^{j\theta}, z) \leq c\} \tag{9.5}$$

Theory of the OPT problem (see [HMar]) implies that the existence and other properties of solutions to OPT are closely related to properties of the sublevel sets of the performance function. Usually "well behaved" sublevel sets correspond to solutions one can consider "nice". A well behaved sublevel set has the following characteristics:

- bounded: the set is contained in some disk with finite radius
- connected: two points in $S_\theta(c)$ can be joined by an arc that lies totally in $S_\theta(c)$
- simply connected: no "holes", i.e., the set complementary to $S_\theta(c)$ is connected
- The gradient of $\Gamma(e^{j\theta}, z)$ in $z = x + jy$ does not vanish when z is in the boundary of $S_\theta(c)$

Other desirable properties are:

- convexity: two points in $S_\theta(c)$ can be joined by a segment that lies totally in $S_\theta(c)$

- smooth boundary
- smooth dependence on θ

Example 1. For the performance function $\Gamma(e^{j\theta}, z) = |1/(e^{j\theta} + 0.5) - z|^2$ the sublevel set $S_\theta(c)$ is a disk, centered at $1/(e^{j\theta} + 0.5)$ with radius $\sqrt{c}$.

Example 2. If $\Gamma(e^{j\theta}, z) = (2+\cos(\theta))|1/(e^{j\theta}+0.5)-z|^2$, then the sublevel set $S_\theta(c)$ is a disk, centered at $1/(e^{j\theta} + 0.5)$ with radius $\sqrt{c/(2+\cos(\theta))}$.

Example 3. Suppose that two continuous functions $k(e^{j\theta})$ and $w(e^{j\theta})$ are given, and that $w(e^{j\theta}) > 0$ for all θ. With these functions we build the performance

$$\Gamma(e^{j\theta}, z) = w(e^{j\theta}) \mid k(e^{j\theta}) - z \mid^2 \tag{9.6}$$

This performance function appears in control applications and it is quite informative to study. The sublevel sets $S_\theta(c)$ of Γ are disks with center $k(e^{j\theta})$ and radius $(c/w(e^{j\theta}))^{1/2}$. Thus $S_\theta(c)$ is connected and simply connected, and it is bounded since the radius is finite. The sets S_θ are also convex and have smooth boundary. Since $k(e^{j\theta})$ and $w(e^{j\theta})$ are continuous, we have that $S_\theta(c)$ varies continuously with θ. If $k(e^{j\theta})$ and $w(e^{j\theta})$ are smooth, then $S_\theta(c)$ varies smoothly in θ for fixed c. Performance functions with sublevel sets that are disks are called *quasicircular*. The following is an example.

Example 4. Suppose that $w_1(e^{j\theta})$ and $w_2(e^{j\theta})$ are given positive functions, and

$$\Gamma(e^{j\theta}, z) = w_1(e^{j\theta}) \mid z \mid^2 + w_2(e^{j\theta}) \mid 1 - z \mid^2 \tag{9.7}$$

This performance function arises in H^∞ -control literature, where it is called *mixed sensitivity*. It is quasicircular, a fact that we now verify. Use the formula $\mid a - b \mid^2 = \mid a \mid^2 - 2\mathrm{Re}\,(\overline{a}\, b) + \mid b \mid^2$ to rewrite (9.7) as

$$\Gamma(e^{j\theta}, z) = (w_1(e^{j\theta}) + w_2(e^{j\theta})) \mid z \mid^2 - 2\mathrm{Re}\,(w_2(e^{j\theta})z) + w_2(e^{j\theta}) \tag{9.8}$$

Collecting terms and completing the square we get

$$\Gamma(e^{j\theta}, z) = (w_1(e^{j\theta}) + w_2(e^{j\theta})) \mid z - \frac{w_2}{w_1(e^{j\theta})+w_2(e^{j\theta})} \mid^2 + \frac{w_1(e^{j\theta})\, w_2(e^{j\theta})}{w_1(e^{j\theta})+w_2(e^{j\theta})} \tag{9.9}$$

Therefore $S_\theta(c)$ is a disk with center $w_2(e^{j\theta})/(w_1(e^{j\theta}) + w_2(e^{j\theta}))$ and radius

$$\frac{1}{w_1(e^{j\theta}) + w_2(e^{j\theta})}\,(c - \frac{w_1(e^{j\theta})\, w_2(e^{j\theta})}{w_1(e^{j\theta}) + w_2(e^{j\theta})}),$$

and Γ is quasicircular.

Example 5. If $w_1(e^{j\theta})$ and $w_2(e^{j\theta})$ are given positive functions, and

$$\Gamma(e^{j\theta}, z) = w_1(e^{j\theta})\ \left(\mathrm{Re}(\ 1/e^{j\theta} - z\ )\right)^2\ +\ w_2(e^{j\theta})\ \left(\mathrm{Im}(\ 1/e^{j\theta} - z\ )\right)^2 \tag{9.10}$$

then $S_\theta(c)$ is an ellipse centered at $1/e^{j\theta}$ and semiaxes that vary with θ.

Example 6. Set

$$\Gamma(e^{j\theta}, z) = |\ 1\ +\ (1/e^{j\theta}\ -\ z)^2\ |^2 \tag{9.11}$$

Then $S_\theta(c)$ is a lemniscate centered at $1/e^{j\theta}$. For small c the sublevel sets are convex, for c slightly smaller than 1 the sublevel sets are non-convex, and for $c > 1$ the sublevel sets are no longer connected sets.

Example 7. Let $w_1(e^{j\theta})$ and $w_2(e^{j\theta})$ be given positive functions, and

$$\Gamma(e^{j\theta}, z) = \max \{\ w_1(e^{j\theta})|\ 1/e^{j\theta} - z\ |\ ,\ w_2(e^{j\theta})|\ \frac{1}{e^{j\theta} - 0.5} - z\ |\ \} \tag{9.12}$$

Then $S_\theta(c)$ is the intersection of two disks centered at $1/e^{j\theta}$ and at $\frac{1}{e^{j\theta}-0.5}$, with radii $c/w_1(e^{j\theta})$ and $c/w_2(e^{j\theta})$.

Example 8. Let $w_1(e^{j\theta})$ and $w_2(e^{j\theta})$ be given positive functions, and

$$\Gamma(e^{j\theta}, z) = w_1(e^{j\theta})\mathrm{Re}(\ 1/e^{j\theta} - z\ )\ +\ w_2(e^{j\theta})\mathrm{Im}(\ \frac{1}{e^{j\theta} - 0.5} - z\ ) \tag{9.13}$$

Then $S_\theta(c)$ is a rhombus centered at $1/e^{j\theta}$.

Chapter 10

Solutions to OPT

Suppose that we are handed a candidate f for a solution to an OPT problem and asked if it is a solution. The main result in this chapter (Theorem 1) provides such a test.

10.1 Complex partial derivatives

Since

$$x = \frac{1}{2}(z + \overline{z})$$

$$y = \frac{1}{2j}(z - \overline{z})$$

we can change the variables in a function g of x and y and write g as a function of the variables z and $\overline{z}$. For example,

$$\begin{aligned} g &= x^2 + y^2 && \Rightarrow \quad g = z\overline{z} \\ g &= x^2 - y^2 + 2xyj && \Rightarrow \quad g = z^2 \\ g &= y - x^2 j && \Rightarrow \quad g = \tfrac{1}{2j}(z - \overline{z}) - (\tfrac{z+\overline{z}}{2})^2 j \end{aligned} \tag{10.1}$$

Partial derivatives with respect to the variables z and $\overline{z}$ are defined by the formulas

$$\begin{aligned} \tfrac{\partial}{\partial z} &= \tfrac{1}{2}(\tfrac{\partial}{\partial x} - j\tfrac{\partial}{\partial y}) \\ \tfrac{\partial}{\partial \overline{z}} &= \tfrac{1}{2j}(\tfrac{\partial}{\partial x} + j\tfrac{\partial}{\partial y}) \end{aligned} \tag{10.2}$$

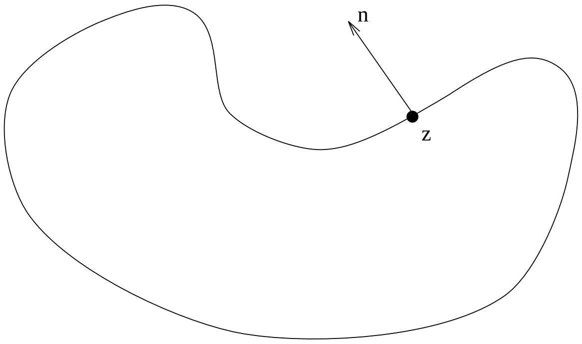

Figure 10.1: The sublevel set $S_\theta(c)$ and the gradient of a real valued function $g(z)$.

The partial derivatives with respect to z and to $\overline{z}$ satisfy the standard differentiation rules for functions of one real variable. Some examples are

$$\begin{aligned} \tfrac{\partial}{\partial z}\left(\mid z \mid^2\right) &= \tfrac{\partial}{\partial z}\left(z\overline{z}\right) = \overline{z} \\ \tfrac{\partial}{\partial z}\left(z^2\right) &= 2z \\ \tfrac{\partial}{\partial \overline{z}}\left(z^2\right) &= 0 \\ \tfrac{\partial}{\partial z}\left(\tfrac{1}{2j}(z-\overline{z}) - (\tfrac{z+\overline{z})}{2})^2 j\right) &= \tfrac{1}{2j} - (z+\overline{z}) \end{aligned} \tag{10.3}$$

It is well known that complex functions $g = u(x,y) + jv(x,y)$ are analytic as a function of $z = x + jy$ if and only if u and v are harmonic functions that satisfy the Cauchy-Riemann equations (c.f. [**A**]). On the other hand it can be shown that the Cauchy-Riemann equations can be written as $\frac{\partial}{\partial \overline{z}} g = 0$. Thus the latter equation is equivalent to saying that g is analytic.

A consequence of the definition of $\partial / \partial \overline{z}$ is that if g is a *real valued* function of z then $\frac{\partial g}{\partial \overline{z}} g$ is precisely the gradient of g in the real variables x and y. Thus we have that (abusing notation),

$$\frac{\partial g}{\partial \bar{z}}(z) = \overline{\frac{\partial g}{\partial z}}(z) = \nabla_{x,y} g(x+iy) = (\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y})$$

See Fig. 10.1.

Although we write functions $\Gamma(e^{j\theta}, z)$ as dependent on z only, actually they also depend on $\overline{z}$. For convenience the variable $\overline{z}$ is not made explicit. For example,

$$\Gamma(e^{j\theta}, z) = \left|\ 1/e^{j\theta} - z\right|^2 \tag{10.4}$$

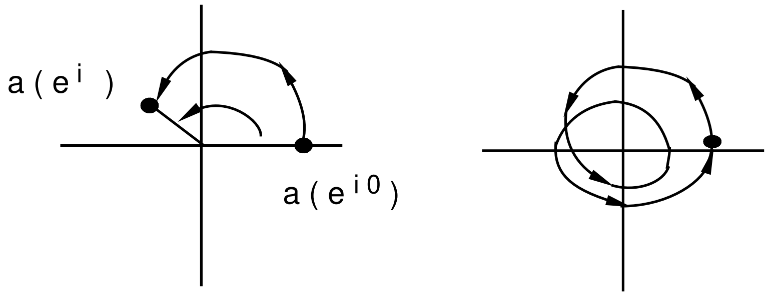

Figure 10.2: The winding number of a function about 0.

is a function of *both* z and $\overline{z}$:

$$\Gamma(e^{j\theta}, z) = (\ 1/e^{j\theta} - z\ )\ \overline{(\ 1/e^{j\theta} - z\ )} = (\ 1/e^{j\theta} - z)\ (\ e^{j\theta} - \overline{z}) \qquad (10.5)$$

In particular we have that

$$\frac{\partial \Gamma}{\partial z}(e^{j\theta}, z) = (-1)\ \overline{(\ 1/e^{j\theta} - z\ )} \qquad (10.6)$$

10.2 Winding number

Another concept we need to introduce is *winding number* of a curve with respect to a point. If $a : \partial D \rightarrow C$ is continuous and never 0, then as θ varies from 0 to 2π there is always a well defined *phase* (or angle) determined by $a(e^{j\theta})$ and the positive semiaxis. Furthermore we can choose the phase to vary continuously with θ, provided we allow for angles to take any real value. This phase function is called a *continuous argument* (c.f. [**A**]).

Since $a(e^0) = a(e^{2\pi i})$, the increment in the value of any continuous argument from $\theta = 0$ to $\theta = 2\pi$ is an integer multiple of 2π, say $2\pi n$ (see Fig. 10.2). The number n is called the "winding number of $a(\cdot)$ with respect to 0, and is denoted wind (a ; 0).

10.3 Main result on the characterization of solutions

The characterization of scalar valued solutions to OPT is

Theorem 13 *Suppose a given performance function* $\Gamma(e^{j\theta}, z)$ *is of class* C^3. *Let* f^* *be a function in* A *for which the function* $a(e^{j\theta}) \stackrel{\Delta}{=} \frac{\partial \Gamma}{\partial z}(e^{j\theta}, f^*(e^{j\theta}))$ *never vanishes. Then* f^* *is a strict local minimizer to OPT if and only if* *(i)* $\Gamma(e^{j\theta}, f^*(e^{j\theta}))$ *is constant in* θ.

(ii) wind (a ; 0) > 0.

The following is a very nice geometric interpretation of (i) and (ii) of Theorem 13. Suppose that γ^* is the optimal value of OPT. The sublevel sets

$$S_\theta(c) = \{z \in C : \Gamma(e^{j\theta}, z) \leq c\}$$

at $c = \gamma^*$ can be represented as a solid in 3D as shown in Fig. 10.3. In this figure the variable θ is represented in one of the axes, and at each constant θ the plane perpendicular to the θ axis represents the complex plane containing the sublevel set S_θ.

Condition (i) in Theorem 13 implies that $f^*(e^{j\theta})$ is always on the boundary of the set $S_\theta(\gamma^*)$. Thus as θ varies from 0 to 2π, the pair $(\theta, f(e^{j\theta}))$ describes a curve on the boundary of the solid shown in Fig. 10.3. In other words, if the curve $(e^{j\theta}, f^*(e^{j\theta}))$ dips inside the solid, or if it goes out of the solid, then f^* does not satisfy condition (i).

To see what (ii) means, consider the function $\overline{a(e^{j\theta})} = \overline{\frac{\partial\Gamma}{\partial z}(e^{j\theta}, f^*(e^{j\theta}))}$. At each θ this represents an exterior normal to the set $S_\theta(\gamma^*) \subset C$ at the point $f^*(e^{j\theta})$. As θ varies from 0 to 2π, the function $\overline{a}$ winds about the θ axis. Condition (ii) of the theorem says that if f^* is an optimizer, then $\bar{a}$ winds at least once in the clockwise sense.

Example 1. The constant function $f(e^{j\theta}) = 0$ is a solution to OPT when the performance is $\Gamma(e^{j\theta}, z) = |1/e^{j\theta} - z|^2$. Indeed, $\Gamma(e^{j\theta}, 0) = 1$ and $\frac{\partial\Gamma}{\partial z}(e^{j\theta}, 0) = -e^{j\theta}$, so conditions (i) and (ii) in Theorem 1 are satisfied.

Exercise. Let α and β be positive constants. Consider the performance

$$\Gamma(e^{j\theta}, z) = \alpha\ \left(\mathrm{Re}(\ 1/e^{j\theta} - z\ )\right)^2 \ +\ \beta\ \left(\mathrm{Im}(\ 1/e^{j\theta} - z\ )\right)^2 \qquad (10.7)$$

Prove that for a suitable c_0, the function $f(e^{j\theta}) = c_0 e^{j\theta}$ is a solution to OPT.

10.4 Properties of solutions

A knowledge of the fundamental qualitative properties of solutions is very useful to someone person who is using a computer program to solve problems like OPT. Then when the program produces odd answers the user can have an idea of what is happening or at least eliminate some possibilities.

Three important questions from the qualitative point of view are:

- Does OPT have a solution or merely an optimizing sequence?

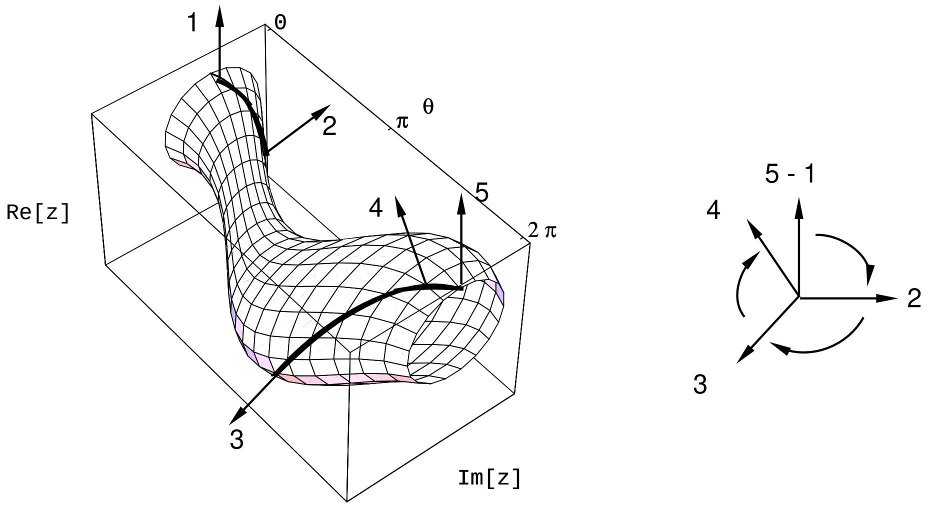

Figure 10.3: The optimal performance surface. In (a), the curve $(e^{j\theta}, f^*(e^{j\theta}))$ lies on the boundary of the solid. As θ increases from 0 to 2π, the arrow representing a normal to the sublevel set makes a complete clockwise revolution with respect to the θ axis, as indicated in (b).

- If a solution exists is it unique?
- If a solution exists is it smooth?

In the present case of scalar valued functions f, the OPT problem is incredibly well behaved (even when the performance function Γ is not convex in z). Solutions exist in A and are unique! Solutions are smooth (infinitely differentiable) if the performance function Γ is smooth. A discussion of existence and uniqueness of solutions in the scalar case can be found in **[HMar]**.

The sublevel sets

$$S_\theta(c) = \{z : \Gamma(e^{j\theta}, z) \leq c\}$$

are basic to the study of qualitative properties of OPT. The *standard assumption* that we make on sublevel sets is

Standard Assumption (SA): $S_\theta(c)$ is connected and simply connected, and as θ varies it is uniformly bounded and has area uniformly bounded away from 0. Furthermore, the gradient of $\Gamma(e^{j\theta}, z)$ in $z = x + jy$ is not zero for all z in the boundary of $S_\theta(c)$.

Theorem 14 *Let* $\Gamma(e^{j\theta}, z)$ *be at least once continuously differentiable with sublevel sets that satisfy (SA). Then solutions* $f^* \in A$ *to OPT exist and are unique. If* Γ *is infinitely differentiable, then* f^* *is infinitely differentiable on the circle.*

One consequence of Theorem 14 is that *global minimizers* and *local minimizers* to OPT are the same. In particular, if one calculates a solution using a method for finding local solutions, then the solution is in fact the global solution.

Also one should not take smoothness for granted. In the very natural and important problem of optimizing H^∞ performance over stable compensators, one finds that the optimal closed loop transfer function is almost never continuous [??]. This notion originated in circuit theory and there it is called the *spiked gain principle* **[H3]**.

The results cited in this section extend to optimization problems where the performance is not necessarily real symmetric, and the optimization is performed over analytic functions that are not necessarily ' “real”, i.e., satisfy $f(e^{-j\theta}) = \bar{f}(e^{j\theta})$.

Chapter 11

Facts about analytic functions

Some basic properties of analytic functions are presented in this section. Section 11.1 presents a result on approximation of functions on an arc of the circle. Spectral factorization is discussed in Section 11.2, and an analog result on phase of functions is presented in Section 11.3. The chapter is closed with a discussion in Section 11.4 of an important design principle.

11.1 Any function on an arc is a small perturbation of a function analytic on the Disk.

Theorem 15 *Let P be a closed subset of the unit circle ∂D which does not equal the whole unit circle. If g is any continuous function on P, then there is a sequence of rational functions $r_k \in A$ which approximate g on P. That is,*

$$\sup_{e^{j\theta} \in P} |r_k(e^{j\theta}) - g(e^{j\theta})| \to 0 \qquad \text{as } k \to \infty$$

Moreover (Runge), if $\mathcal{R}$ is any subset of the complex plane C which does not intersect the closed unit disk, then one can choose each r_k to have its poles in $\mathcal{R}$.

Proof: One can use a suitable (analytic) linear fractional transformation $v = l(\zeta)$ to map the unit disk to the unit disk, and the set P to a set P_1 that is a subset of the open disk $D(1,1)$ centered at 1 with radius 1. Then to approximate g on P by analytic functions is equivalent to approximating the function $g_1 = g \circ l^{-1}$ on P_1 by analytic functions of v: if f is a function analytic on the disk such that

$$|g_1(e^{j\theta}) - f(e^{j\theta})| < \epsilon \qquad \text{for } e^{j\theta} \in P_1$$

then clearly $f \circ l^{-1}$ is analytic and

$$|g(e^{j\theta}) - f(l^{-1}(e^{j\theta}))| < \epsilon \qquad \text{for } e^{j\theta} \in P$$

Thus we may assume that P is a set contained in the disk $D(1,1)$, and we make this assumption in the rest of the proof.

As a first case consider a function $g(\zeta) = \sum_{\ell=-n}^{n} c_\ell \zeta^\ell$ where $n > 0$ is fixed. The function $g(\zeta)$ can be expanded in a power series of $\zeta - 1$ that is uniformly and absolutely convergent on compact subsets of $D(1,1)$. By taking as many terms as necessary in this power series, we can approximate $g(e^{j\theta})$ by analytic polynomials of $e^{j\theta}$.

The same holds when g is a continuous function on P, since such functions are approximable by expressions of the form $\sum_{\ell=-n}^{n} c_\ell (e^{j\theta})^\ell$ (Weierstrass approximation Theorem).

Therefore g is approximable by analytic functions on P. Finally, Runge's Theorem **[R]** guarantees that any analytic function on the disk that is continuous on the closed unit disk can be uniformly approximated on the circle by rational functions whose poles are in any specified set that does not intersect the closed disk. ■

How badly behaved are the functions r_k in Theorem 15 *outside* the arc R? The answer is given by the following theorem:

Theorem 16 *Let P be a closed subset of the unit circle ∂D which does not equal the whole unit circle. If g is any continuous function on P, then either g extends to a bounded analytic function on the disk D, or, for any sequence $\{f_n\} \subset A$ such that*

$$\sup_{e^{j\theta} \in P} |f_n(e^{j\theta}) - g(e^{j\theta})| \to 0 \qquad \text{as } n \to \infty,$$

the sequence must satisfy $\sup_\theta |f_n(e^{j\theta})| \to \infty$ as $n \to \infty$.

Since the proof of this theorem requires more advanced mathematics, we provide only a sketch here. If $\sup_\theta |f_n(e^{j\theta})| \le k_0$ for some constant k_0 and all n, then $\{f_n\}$ is a normal family, c.f. **[A]**. Therefore it has a subsequence which is convergent uniformly on compact subsets of the disk to a function $f_0 \in H^\infty$. Moreover, $|f_n(e^{j\theta}) - f_0(e^{j\theta})| \to 0$ as $n \to \infty$, for all $e^{j\theta} \in P$. Then $g(e^{j\theta}) = f_0(e^{j\theta})$ for $e^{j\theta} \in P$, so f_0 is an extension of g analytic on the disk.

11.2 Spectral factorization: analytic functions with prescribed amplitude

Theorem 17 *Given a differentiable function* $r(e^{j\theta}) > 0$*, there exists a unique function* $f \in A$ *that is never 0 on the closed disk* D^- *and such that factors* r *in the sense*

$$r(e^{j\theta}) = f(e^{j\theta})\overline{f(e^{j\theta})}$$

Moreover, if r *is a polynomial or rational, then so is* f*.*

Proof: We give a proof for the case when r is a polynomial. Suppose that r is given by

$$r(e^{j\theta}) = \sum_{n=-N}^{n=N} c_n (e^{j\theta})^n.$$

Since r is real valued, we have that $r(e^{j\theta}) = \overline{r(e^{j\theta})}$, and from this it follows that $c_{-n} = \overline{c_n}$. Now $r(\zeta)$ is a rational function of $\zeta \in C$, with poles located at $\zeta = 0$, and no zeros on the unit circle. If $\zeta_0 \in C$ is such that $r(\zeta_0) = 0$, we have that

$$\begin{aligned} 0 &= \overline{r(\zeta_0)} \\ &= \textstyle\sum_{n=-N}^{n=N} \overline{c_n}(\overline{\zeta_0})^n \\ &= \textstyle\sum_{n=-N}^{n=N} c_{-n}(\overline{\zeta_0})^n \\ &= \textstyle\sum_{n=-N}^{n=N} c_{-n}(\frac{\zeta_0}{|\zeta_0|^2})^{-n} \\ &= r(\frac{\zeta_0}{|\zeta_0|^2}) \end{aligned} \tag{11.1}$$

We have proved that if ζ is a zero of r, then $\zeta_0^* = \frac{\zeta_0}{|\zeta_0|^2}$ is also a zero. Thus the polynomial $(\zeta - \zeta_0)(\zeta - \zeta_0^*)$ is a factor of the numerator of

$$r(\zeta) = \frac{c_{-N} + c_{-N+1}\zeta + \cdots + c_{N-1}\zeta^{2N-1} + c_N\zeta^{2N}}{\zeta^N}$$

which (as the reader can easily prove), upon division, yields a polynomial exhibiting the same type of symmetric arrangement of the coefficients. One can carry this process until r is completely factored in the following form:

$$r(\zeta) = \frac{(\zeta - \zeta_1)\cdots(\zeta - \zeta_N)(\zeta - \zeta_1^*)\cdots(\zeta - \zeta_N^*)}{\zeta^N} \tag{11.2}$$

Since $|\zeta| < 1$ if and only if $|\zeta^*| > 1$, we may label $\zeta_1, \ldots, \zeta_N$ as the roots of $r(\zeta)$ located in the unit disk so that $\zeta_\ell^*, \ldots, \zeta_N^*$ are outside the unit disk. Now set

$$f(\zeta) = (\zeta - \zeta_1) \cdots (\zeta - \zeta_N).$$

Note that the relation

$$e^{j\theta} - \zeta_0^* = e^{j\theta} - \frac{\zeta_0}{|\zeta_0|^2} = \frac{\zeta_0}{|\zeta_0|^2}(-e^{j\theta})(\overline{e^{j\theta}} - \overline{\zeta})$$

implies that $r(e^{j\theta}) = f(e^{j\theta})\overline{f(e^{j\theta})}$. This proves the theorem in the case when r is a polynomial.

11.3 Analytic functions with prescribed phase

Theorem 18 *Let b be a complex valued, differentiable function of $e^{j\theta}$ that never equals 0. Then, there is a smooth function $f \in A$ whose phase equals the phase of b at every $e^{j\theta}$ if and only if* wind $(b; 0) = n \geq 0$. *Furthermore, the function f has no zeros in the disk if and only if $n = 0$, and in this case f is unique up to a constant. We call this function f the phase (analytic) factor of b.*

Proof: This may be beyond many students so not everyone will want to spend much time on it. Suppose that b is a given differentiable function on the circle with wind $(b; 0) = n = 0$, and let $\log(b) = w + iv$. Since v is a continuous function on the circle, it has an associated Fourier series expansion

$$v(e^{j\theta}) = \sum_{\ell=-\infty}^{\infty} \hat{v}_\ell e^{j\ell\theta} \tag{11.3}$$

where (since v is real valued) $\overline{\hat{v}_\ell} = \hat{v}_{-\ell}$. Now let u be the function on the circle whose Fourier coefficients are given by $\hat{u}_\ell = -i\hat{v}_\ell$ if $\ell < 0$, $\hat{u}_\ell = i\hat{v}_\ell$ if $\ell > 0$, and $\hat{u}_0 = 0$. Set $f(e^{j\theta}) = u(e^{j\theta}) + iv(e^{j\theta})$. A calculation yields

$$g(e^{j\theta}) := u(e^{j\theta}) + iv(e^{j\theta}) = \hat{v}_0 + \sum_{\ell=1}^{\infty} 2\hat{v}_\ell e^{j\ell\theta} \tag{11.4}$$

Since b is differentiable, then v is differentiable too and then **[Hof]** the function u is a continuous function of $e^{j\theta}$. Hence g is a continuous function on the circle. In the theory of Hardy Spaces it is shown that (11.4) can be used to define a continuous extension $\tilde{g}$ of g to the closed unit disk, and analytic inside. The Taylor coefficients in the expansion of $\tilde{g}$ about 0 are precisely the Fourier coefficients of g:

$$g(\zeta) = \hat{v_0} + \sum_{\ell=1}^{\infty} 2\hat{v_\ell}(\zeta)^\ell \tag{11.5}$$

Finally[1], set $f(\zeta) = e^{g(\zeta)}$. Clearly f has no zeros in D.

To prove uniqueness, note that if $\tilde{f}$ is another analytic function whose phase on the circle is the phase of b, then $\tilde{f}/f$ is analytic on D since f has no zeros there. Also, this ratio is real valued. But this is not possible unless the ratio is a constant.

Now, if $n > 0$ then consider the function $b_1(e^{j\theta}) = b(e^{j\theta})/(e^{j\theta})^n$. By the above discussion, there is a function f analytic on the disk whose phase on the circle is the phase of b_1. But then the phase of $b(e^{j\theta})$ is equal to the phase of $f(e^{j\theta})(e^{j\theta})^n$, which is analytic on the disk D.

We just proved one direction of the Theorem. The other direction is easy: if f is an analytic function with the same phase as the phase of b, then $n =$ number of zeros of f inside the disk ≥ 0. In particular, if f has no zeros, then $n = 0$. ■

We finish the subsection by mentioning a basic result about rational functions (see **[A]**)

Theorem 19 *If $r(\zeta)$ is a rational function that has no poles or zeros on the unit circle $|\zeta| = 1$, then*

$$\text{wind } (\ r\ ;\ 0\ ) = z_r - p_r$$

where z_r = number of zeros of r inside the unit disk, and p_r = number of poles of r inside the unit disk.

---

[1] Another approach to proving the Theorem is to construct the solution to Laplace's equation

$$\begin{aligned} \Delta v(x,y) &= 0, \qquad (x,y) \in D \\ v|_{\partial D} &= v_0 \end{aligned} \tag{11.6}$$

for example using Poisson's formula. Now a harmonic function $v^*(x,y)$ can be determined so that $g(x+iy) = v^*(x,y) + iv(x,y)$ satisfies the Cauchy-Riemann equations on D. The function $f = e^g$ is analytic on the disk D, and its phase on the circle is v, i.e., the phase of b.

11.4 The fundamental mistake of H^∞ control and other stories.

This section is not essential to the proofs which follow in the rest of the book, however, the discussion can help the student reinforce the ideas in section (11.1). It also states a simple principle which every control designer should be familiar with.

Any frequency domain control problem which has been correctly translated to a math problem suitable for computer optimization has the property that at each frequency there is a constraint imposed by the specifications. People who violate this make

The fundamental mistake of H^∞-Control

> *"I am only interested in a frequency band 'B', so I will not put specs outside of the band B."*

The price of making this mistake is high because, upon running an optimization program with such specs as input, one finds that the closed loop transfer function T has huge magnitude off of B. Amusingly enough T will meet the specs perfectly on B. In short, one obtains a solution so ridiculous that it is not even a good basis for trail and error.

Why are the remarks we just made true? Well, they are an immediate consequence of Theorem 15 and of Theorem 16; indeed the remarks are just a rephrasing of the mentioned results only now in the context of control.

A wholly different consequence of Runge's Theorem is that *pole placement*, a valuable technique in control design, has a shaky foundation which (in the authors opinion) has not been carefully analyzed. The basic idea behind pole placement is to set the poles of the closed loop transfer function at some prescribed distance from the $j\omega$ axis to achieve an objective, like stabilizing the system without making it too sluggish. While the technique is effective in particular situations where there is a lot of human intervention and time for trial and error, we maintain that the method is badly codified so it cannot be used algorithmically. This is a consequence of Runge's Theorem (though not of Theorem 15). Runge's Theorem implies that any function in A_{RHP} and smooth on the $j\omega$ axis (even at $\omega = \infty$) can be approximated by r_k's exactly as in Theorem 15. That is, the hypotheses of this theorem are different, but the conclusion is the same. This of course implies that if we have a proposed design producing a closed loop transfer function T_0 that is rational with poles at locations $\{p_j\}$, then we can find T_ε such that $||T - T_\varepsilon||_\infty \leq \varepsilon$ and with poles about any place we want. The number of poles of T_ε might be greater than that of T_0. This and other constraints that one might impose in order to make pole placement a systematic rather than trial and error subject have not been analyzed.

The moral of the pole placement story is different from the tale of the fundamental mistake. Here we are saying that pole placement can be a viable method, but it is a ripe area for more research. To the contrary the fundamental mistake should never be made.

Chapter 12

Proof of the main result

An important tool used in the proof of Therorem 13 is the Taylor expansion of smooth real valued functions of $z \in C$ in powers of z. This is applied to performance functions in Section 12.1. The actual proof of the theorem begins with the necessity of conditions (i) and (ii) of the Theorem. This is the contents of Section 12.2 and Section 12.3 respectively. The proof of the theorem is completed with the discussion of sufficiency of (i) and (ii) in Section 12.4.

12.1 Taylor expansions and performance functions

By using relation (10.2) the reader can verify that if $G(z)$ is a real valued and smooth function, then the following expansion holds:

$$G(z) = G(0) + 2\mathrm{Re}\,(\frac{\partial G}{\partial z}(0)z) + \mathcal{O}(|z|^2) \qquad (12.1)$$

To use this in the proof of the Theorem, we consider the function $G(z) = \Gamma(e^{j\theta}, f^*(e^{j\theta}) + z)$ (for now $e^{j\theta}$ is fixed), and for ease of notation we write

$$a(e^{j\theta}) = \frac{\partial \Gamma}{\partial z}(e^{j\theta}, f^*(e^{j\theta})) \qquad (12.2)$$

The Taylor expansion can now be restated as

$$\Gamma(e^{j\theta}, f^*(e^{j\theta}) + z) = \Gamma(e^{j\theta}, f^*(e^{j\theta})) + 2\mathrm{Re}a(e^{j\theta})z + \mathcal{O}(e^{j\theta}, |z|^2) \quad (12.3)$$

Figure 12.1: The function $g(e^{j\theta})$ dips below $\gamma - \delta$.

The $\mathcal{O}$ -term in (12.3) represents a function that goes to 0 as least as fast as a constant multiple of $|z|^2$ when $z \longrightarrow 0$. In other words when $|z|$ is small,

$$\Gamma(e^{j\theta}, f^*(e^{j\theta}) + z) \approx \Gamma(e^{j\theta}, f^*(e^{j\theta}) + 2\mathrm{Re}(a(e^{j\theta})z) \qquad (12.4)$$

Of course the $\mathcal{O}$ -term in equation (12.3) depends on θ, but it is possible to obtain an inequality that is valid for all $e^{j\theta}$: the hypotheses on smoothness of Γ imply that there exist positive constants c_0 and δ_0 such that the following relation holds:

$$-c_0|z|^2 \leq \mathcal{O}(e^{j\theta}, |z|^2) \leq c_0|z|^2, \quad \forall\ e^{j\theta}, \ |z| < \delta_0 \qquad (12.5)$$

We leave the proof of (12.5) to the reader.

12.2 Solutions make the performance flat

Suppose that f^* is a solution to OPT for which $\Gamma(e^{j\theta}, f^*(e^{j\theta}))$ is not constant. By the assumptions, there is a θ-interval τ on which $\Gamma(e^{j\theta}, f^*(e^{j\theta}))$ dips below $\gamma^* - \delta_1$ (see Fig. 12.1). To obtain a contradiction we will produce a function $f \in A$ with formula $f = f^* + \epsilon h$ and with the property that $\sup_\theta \Gamma(e^{j\theta}, f(e^{j\theta})) < \sup_\theta \Gamma(e^{j\theta}, f^*(e^{j\theta}))$. To construct and analyze f we will use the set τ.

By Theorem 15 we can select $h \in A$ with phase that approximates the phase of $-\overline{a(e^{j\theta})}$ on τ^c, the complement of τ. That is,

$$\begin{array}{rcl} \text{phase}\ (a(e^{j\theta})h(e^{j\theta})) & \approx & \text{phase}\ (a(e^{j\theta})(-\overline{a(e^{j\theta})})) \\ & = & \text{phase}\ (-|a(e^{j\theta})|^2) \quad = \quad -180^\circ \quad \forall\theta \in \tau^c \end{array}$$

This uses only a small part of the strength of Theorem 15 in that we could have chosen h to approximate the whole function $-\overline{a(e^{j\theta})}$ on τ^c. Thus, there exists a constant $\delta_2 > 0$ such that

$$\text{Re } a(e^{j\theta})h(e^{j\theta}) < -\delta_2 < 0 \qquad \forall e^{j\theta} \in \tau^c \tag{12.6}$$

We use the Taylor expansion of Γ to see that for $\epsilon > 0$ small enough[1] the function $f = f^* + \epsilon h$ satisfies

$$\sup_\theta \Gamma(e^{j\theta}, f(e^{j\theta})) < \sup_\theta \Gamma(e^{j\theta}, f^*(e^{j\theta})) - \epsilon\delta_2 \qquad \forall\theta \in \tau^c \tag{12.8}$$

Of course ϵh might be very large on τ, and if ϵ is not small enough, then $\Gamma(e^{j\theta}, f(e^{j\theta}))$ may be larger than γ^* for $\theta \in \tau$. However, we can take ϵ so small that

$$\sup_{\theta\in\tau} |\Gamma(e^{j\theta}, f^*(e^{j\theta}) + \epsilon h(e^{j\theta}))| < \gamma^* - \delta_1/2 \tag{12.9}$$

Then $\Gamma(e^{j\theta}, f(e^{j\theta}) + \epsilon h(e^{j\theta}))$ is uniformly less than γ^* for *all* θ, contradicting the assumption that f^* is optimal.

12.3 Solutions satisfy the winding number condition

We now may assume $\Gamma(e^{j\theta}, f^*(e^{j\theta}))$ is constant $= \gamma^*$. If $m := \text{wind } (\ a \ ; \ 0 \ ) \le 0$, then the function $b(e^{j\theta}) = (e^{j\theta})^m a(e^{j\theta})$ is never 0 and has winding number 0 about 0. Then by Theorem 18 there is a function h in A whose phase approximates the phase of $-\overline{b}$, so that there exists a constant $\delta_2 > 0$ such that

$$2\text{Re}a(e^{j\theta})h(e^{j\theta}) < -\delta_2 < 0 \ \ \forall\theta \tag{12.10}$$

Compare (12.10) with (12.6): the former is valid for all θ and the latter is valid for $\theta \in \tau^c$. We can proceed with steps similar to those that follow (12.6) and with a small enough ϵ we obtain

$$\sup_\theta \Gamma(e^{j\theta}, f^*(e^{j\theta}) + \epsilon h(e^{j\theta})) < \gamma^* - \epsilon\delta_2/2, \ \forall\theta$$

This produces the contradiction

$$\sup_\theta \Gamma(e^{j\theta}, f(e^{j\theta})) < \sup_\theta \Gamma(e^{j\theta}, f^*(e^{j\theta}))$$

---

[1] A precise argument is, choose $\epsilon > 0$ such that $\epsilon c_0 ||h||^2_\infty < -\delta_2$, that is, $c_0||\epsilon h||^2_\infty < -\epsilon\delta_2$. Combining this with inequality (12.5) we have that

$$\begin{aligned} \Gamma(e^{j\theta}, f^*(e^{j\theta}) + \epsilon h(e^{j\theta})) &\le \gamma^* + 2\text{Re}a(e^{j\theta})\epsilon h(e^{j\theta}) + c_0|\epsilon h(e^{j\theta})|^2 \\ &< \gamma^* - 2\delta_2\epsilon + \epsilon\delta_2 = \gamma^* - \epsilon\delta_2, \ \forall\theta \in \tau^c \end{aligned} \tag{12.7}$$

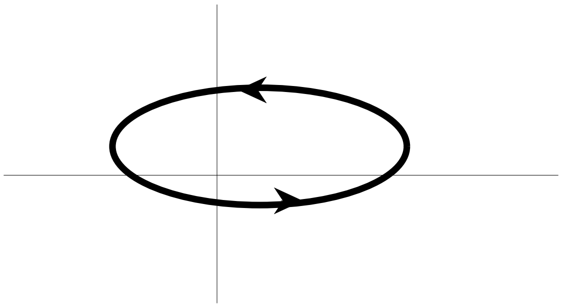

Figure 12.2: The function ah_n winds about 0.

12.4 Flatness and winding number conditions are sufficient

Suppose now that conditions (i) and (ii) hold, but that f^* is not an optimizer. To obtain a contradiction we assume that $\{h_n\}_{n\in N}$ is a sequence in A such that $||h_n||_\infty \to 0$ and

$$\sup_\theta \Gamma(e^{j\theta}, f^*(e^{j\theta}) + h_n(e^{j\theta})) < \sup_\theta \Gamma(e^{j\theta}, f^*(e^{j\theta})) =: \gamma^* \tag{12.11}$$

Before launching into heavy details we give the idea of the proof. Taylor approximation and the flatness condition (i) of Theorem 13 imply that

$$\Gamma(e^{j\theta}, f^*(e^{j\theta}) + h_n(e^{j\theta})) \approx \gamma^* + 2\mathrm{Re}(a(e^{j\theta})h_n(e^{j\theta})). \tag{12.12}$$

The key point is that since h_n is analytic on the disk, then

$$\text{wind}\,(ah_n; 0) = \text{wind}\,(a; 0) + \text{wind}\,(h_n; 0) > 0$$

provided h_n has no zeros on the circle. (If it does, then one can replace it by a function with zeros on the circle that is a small perturbation of h_n). This says that the function ah_n must wind about 0 (see Fig. 12.2), so in particular $\mathrm{Re}(ah_n)$ must be *positive* for some $e^{j\theta_1}$. Hence

$$\Gamma(e^{j\theta_1}, f^*(e^{j\theta_1}) + h_n(e^{j\theta_1})) \approx \gamma^* + 2\mathrm{Re}(a(e^{j\theta_1})h_n(e^{j\theta_1})) > \gamma^*$$

But this contradicts the assumption (12.11).

Now we give a formal proof which includes the second order term in the Taylor expansion. Since $||h_n||_\infty \to 0$ we may assume that $||h_n||_\infty < \delta_0$

without loss of generality, and we have from (12.5) that

$$\begin{aligned} -c_0|h_n(e^{j\theta})|^2 \quad &\leq \quad \Gamma(e^{j\theta}, f^*(e^{j\theta}) + h_n(e^{j\theta})) - \gamma^* - 2\mathrm{Re}(a(e^{j\theta})h_n(e^{j\theta})) \\ &\leq c_0 \quad |h_n(e^{j\theta})|^2, \ \ \forall\ e^{j\theta} \end{aligned} \tag{12.13}$$

Hence

$$\begin{aligned} -c_0|h(e^{j\theta})|^2 + \gamma^* \quad &+ \quad 2\mathrm{Re}(a(e^{j\theta})h_n(e^{j\theta})) \\ &\leq \quad \Gamma(e^{j\theta}, f(e^{j\theta}) + h_n(e^{j\theta})) \leq \gamma^* \quad \forall e^{j\theta} \end{aligned} \tag{12.14}$$

i.e.,

$$-\, c_0|h(e^{j\theta})|^2 + 2\mathrm{Re} a(e^{j\theta})h_n(e^{j\theta}) < 0 \ \ \forall e^{j\theta}. \tag{12.15}$$

Suppose for a moment that for each n there exists θ such that $a(e^{j\theta})h_n(e^{j\theta}) = t_n > 0$. Substituting in (12.15) we get

$$-\frac{c_0}{|a(e^{j\theta})|}t_n^2 + 2t_n < 0 \tag{12.16}$$

i.e.,

$$-\frac{c_0}{|a(e^{j\theta})|}t_n + 2 < 0 \tag{12.17}$$

But since $||h_n||\infty \to 0$, we have that $t_n \to 0$, so the left hand side of 12.17 cannot hold for all n.

To justify the existence of t_n as above, we note that the complex function $H_n(z) = zh_n(z)$ has a zero at $z = 0$. By complex function theory, the function H maps onto some disk centered at 0 so for some θ we have $H_n(e^{j\theta}) > 0$. Since the function $a(e^{j\theta})/e^{j\theta}$ is continuous and never 0, it follows that for some θ,

$$\frac{a(e^{j\theta})}{e^{j\theta}}H_n(e^{j\theta}) = a(e^{j\theta})h_n(e^{j\theta}) = t_n > 0 \tag{12.18}$$

This completes the proof of the Theorem. ∎

Chapter 13

Computer Solutions to OPT

One of the big payoffs of having results such as Theorem 13 available is that one can develop optimality tests. These tests that can be used as a measure of how close a candidate to solution to OPT is to being optimal. This topic is presented in Section 13.1. Section 13.3 describes how Theorem (13) leads to an effective computer algorithm. This requires some basic definitions which are given in Section 13.2.

13.1 Computer diagnostics

An extremely valuable tool in computation, especially with iterative methods, is a collection of *diagnostics* which indicate if the iteration is close to a solution or not close to a solution. The diagnostics which over the years we found effective for OPT are based directly on the general mathematical Theorem 13. This has the advantage that the diagnostics are not algorithm specific, so we recommend them to anyone producing H^∞ software. Indeed we make heavy use of them in *Anopt*, our H^∞ optimization software program.

Suppose that $f \in A$ is a candidate for a solution to OPT. To test Flatness Condition of Theorem 13, we can calculate

$$OT_1(f) := \frac{\sup_\theta \ \Gamma(e^{j\theta}, f(e^{j\theta})) - \inf_\theta \ \Gamma(e^{j\theta}, f(e^{j\theta}))}{\sup_\theta \ \Gamma(e^{j\theta}, f(e^{j\theta}))}$$

To implement [1] condition II, compute

$$OT_2(f) := \text{wind}\ (\ a\ ;\ 0\ ) - 1$$

by evaluating the phase of f on a grid and taking differences of these values on adjacent grid points. At optimum $OT_1(f) = 0$ and $OT_2(f) \geq 0$. Optimization programs for solving OPT should print $OT_1(f)$ and $OT_2(f)$ at each iteration so that the user can monitor how the approximate solutions f are progressing. Of course, if f is near the solution we must have $OT_1(f) \approx 0$ and $OT_2(f) \geq 0$.

To illustrate a practical set of diagnostics we solve the problem OPT when Γ is given by

$$\Gamma(e^{j\theta}, z) = |0.8 + (1/e^{j\theta} + z)^2|^2$$

A computer run produces the numbers shown in Table 1.

Table 1

```
It      Current Value       Optimality Tests      True error
          SupGamma            OT1       OT2

0   3.24            E+00   9.9E-01   0  E+00       ??????
1   1.3456107405895E+00   4.8E-01   0  E+00       ??????
2   1.0242722917577E+00   4.9E-02   0  E+00       ??????
3   1.0100847099506E+00   1.8E-02   0  E+00       ??????
4   1.0005821314497E+00   2. E-03   0  E+00       ??????
```

In our run above, the diagnostic OT1 went from 0.99 down to 0.002 in 4 iterations, while OT2 was zero throughout the iteration (this is not unusual for "scalar" cases). The column "true error" was added to illustrate that, typically, the true error $||f_k - f^*||_\infty$ is of the same order of magnitude than the numbers under OT1 and OT2.

The numbers in Table 1 were obtained with the program Anopt. The actual input used in this example is:

```
<<Anopt';
```

---

[1] Since the winding number of a function can only take integer values, small changes in a function typically do not change the winding number of it with respect to 0. Condition II is automatically satisfied for all functions f that are close (in the norm of the supremum) to a solution f^*!

```
g = Abs[0.8 + (1/e + z[1] )^2 ]^2;

     Anopt[g, 0.01];
```

The reader who wants more details may see the document Anopt.ma in the Appendix.

13.2 Spaces of functions

This section serves as background for Section 13.3 and for the appendices.

Functions $f(z)$ in A have a power series expansion about 0 valid for any z in the unit disk:

$$f(z) = \sum_{\ell=0}^{\ell=\infty} c_\ell z^\ell \tag{13.1}$$

where the coefficients a_ℓ are real numbers. On the other hand, if we now consider $f(e^{j\theta})$ as a function on the circle, it is a continuous function and thus is has a Fourier series expansion,

$$f(e^{j\theta}) = \sum_{\ell=-\infty}^{\ell=\infty} \hat{f}_\ell (e^{j\theta})^\ell \tag{13.2}$$

where the Fourier coefficients $\hat{f}_\ell$ are real numbers. The relationship between (13.1) and (13.2) becomes clear when we formally substitute $e^{j\theta}$for z in (13.1):

$$f(e^{j\theta}) = \sum_{\ell=0}^{\ell=\infty} c_\ell (e^{j\theta})^\ell \tag{13.3}$$

Now, by the theory of Fourier series we know that the representation (13.2) is *unique*. If (13.3) is valid, then we must have that $c_n = \hat{f}_n$ for $n = 0, 1, 2, \ldots$, and that $c_n = 0$ for $n = -1, -2, \ldots$. To prove that (13.3) holds for functions $f \in A$, one can consider first polynomials in z for which this relation is obvious. Then the general result follows from the fact that functions in A are limits of polynomials.

Now suppose that $f(e^{j\theta})$ is any function on the circle that is continuous. If the Fourier expansion of $f(e^{j\theta})$ has the form

$$f(e^{j\theta}) = \sum_{\ell=-\infty}^{\ell=\infty} \hat{f}_\ell (e^{j\theta})^\ell \tag{13.4}$$

we see that we can generate a function $f(z)$ defined on the disk as a power series in z, by formally replacing $e^{j\theta}$ by z in (13.4). It is possible to show that the radius of convergence is 1 (see **[Hof]**). Thus, roughly speaking, functions defined by relation (13.1) and functions defined by relation (13.2) can be identified.

If we drop from the definition of A the requirement that $f(z)$ be continuous on the closed unit disk, then we obtain a normed vector space called the Hardy space H^{∞}. It is obvious that some functions $f(z)$ in H^{∞} have boundary values $f(e^{j\theta})$ (those in A for example). But it is not clear, a priori, that one can talk about boundary values of an arbitrary function $f(z)$ in H^{∞}. There is however a way to define $f(e^{j\theta})$ mathematically with the help of *Measure Theory* (see **[Hof]**).

If $f(e^{j\theta})$ is a *measurable* function $f : \partial D \rightarrow C$, the norms below are well defined and give a nonnegative value or infinity:

$$||f||_p = \{\int_0^{2\pi} |f(e^{j\theta})|^p \frac{d\theta}{2\pi}\}^{1/p} \tag{13.5}$$

and

$$||f||_{\infty} = esssup_{\theta} |f(e^{j\theta})| \tag{13.6}$$

The norms above are used to define the normed vector spaces

$$L^p = \{f : f \text{ is a measurable function on } \partial D \text{ and } ||f||_p < \infty\}, \tag{13.7}$$

which are Banach spaces.

The spaces that occur most often correspond to $p = 1, 2, \infty$. In particular, the following relation holds:

$$||f||_1 \leq ||f||_2 \leq ||f||_{\infty}, \tag{13.8}$$

therefore $L^{\infty} \subset L^2 \subset L^1$.

The space L^2 is particularly important because it is also a *Hilbert space* with inner product

$$< f, g >= \int_0^{2\pi} f(e^{j\theta})\overline{g(e^{j\theta})}\frac{d\theta}{2\pi} \tag{13.9}$$

All functions in L^1 do have an associated Fourier series expansion[2] (13.2). This fact can be used to define the Hardy spaces H^p,

$$H^p = \{f \in L^p : f(e^{j\theta}) = \sum_{\ell=0}^{\ell=\infty} \hat{f}_{\ell} e^{j\ell\theta}\}$$

---

[2] But this expansion may fail to give $f(e^{j\theta})$ for some values of $e^{j\theta}$.

Again, functions in H^p can be thought of as functions defined and analytic on the open disk by means of a Taylor series.

Since H^2 is a closed subspace of L^2 it has an orthogonal complement ${H^2}^\perp$ in L^2. It is easy to check that

$$ {H^2}^\perp = \{ \, f \in L^2 \; : \; f(e^{j\theta}) = \sum_{\ell=-1}^{\ell=-\infty} \hat{f}_\ell \, e^{j\ell\theta} \, \}. $$

If $f = \sum_{\ell=-\infty}^{\ell=\infty} \hat{f}_\ell e^{j\ell\theta}$ is in L^2, we define the operators P^- and P_0^- on L^2 by

$$ \begin{aligned} P^-[\, f \,] \; &= \; P^-[\, \textstyle\sum_{\ell=-\infty}^{\ell=\infty} \hat{f}_\ell \, e^{j\ell\theta} \,] = \; \textstyle\sum_{\ell=-\infty}^{\ell=-1} \hat{f}_\ell \, e^{j\ell\theta} \\ P_0^-[\, f \,] = \; & \; P_0^-[\, \textstyle\sum_{\ell=-\infty}^{\ell=\infty} \hat{f}_\ell \, e^{j\ell\theta}] = \; \textstyle\sum_{\ell=-\infty}^{\ell=0} \hat{f}_\ell \, e^{j\ell\theta} \end{aligned} \tag{13.10} $$

P^- is the orthogonal projection of L^2 onto ${H^2}^\perp$.

13.3 A Numerical Algorithm for solving OPT

Suppose that a function $T : R \rightarrow R$ is given. In calculus, students learn that the points $x \in R$ at which T attains a minimum satisfy $T'(x) = 0$ (i.e., x is a critical point), and that one can determine excellent candidates for x by finding these critical points. We will do exactly this in order to find minimizers to OPT, but in a setting where the element x belongs to function space, and T is a function on this space. Consequently, our strategy is to write down some equations $T'(f) = 0$ that minimizers f^* satisfy, and then solve the equations with an iterative method.

The most common method for solving nonlinear equations $T'(x) = 0$ is Newton's method. It is based on the following simple idea. Say x_k is your current guess at the solution x^*. To update you want to find h so that $T'(x_k + h) = 0$, but since you cannot solve for h, you settle for solving for h in the linearized equation

$$ T(x_k) + T'_{x_k}(h) = 0 \tag{13.11} $$

This equation is linear in h and so it is readily solvable for h provided that the linear map T'_{x_k} is invertible. So one solves (13.11) for h, and then set[3] $x_{k+1} = x_k + h$.

---

[3] Often $x_{k+1} = x_k + t * h$ is used, where $t > 0$ is a parameter chosen according to a prespecified criterion. The process of finding t is called linesearch.

One virtue of Newton's method is that it applies to a wide variety of situations, even when the unknowns belong to function spaces. In particular, it can be used to solve OPT by iteration, and this is the basis for the method presented in this section. We treat the scalar case here, but most of what we say also applies to the case of optimization over vector valued functions f (Chapter 15).

The first thing we do now is to rewrite the winding number condition (ii) of Theorem 13 in a more algebraic form. This makes it more suited to a Newton method solution.

Lemma 2 *In addition to the hypotheses to Theorem 13, assume that the function $a(e^{j\theta})$ is smooth. Then the conditions (i) and (ii) of Theorem 13 are equivalent to:*
There exist $\gamma \in R$, $f^* \in A$, $F \in A$, and λ a smooth positive valued function on the circle such that

$$\begin{aligned} \Gamma(\cdot, f) &= \gamma \\ \lambda^{-1} \tfrac{\partial \Gamma}{\partial z}(\cdot, f) &= e^{j\theta} F \end{aligned} \tag{13.12}$$

Proof. Suppose that f^* is such that conditions (i) and (ii) of Theorem (13) are satisfied. From (i) we immediately get the existence of the constant γ in (13.12). If wind $(a ; 0) = n$, we have that the smooth function $b(e^{j\theta}) = a(e^{j\theta})/(e^{j\theta})^n$ has winding number 0 about 0. By Theorem 18 there exists an analytic function F_1 such that phase(b) = phase(F_1). Another way to describe this relation is

$$\frac{b(e^{j\theta})}{|b(e^{j\theta})|} = \frac{F(e^{j\theta})}{|F_1(e^{j\theta})|}, \qquad \forall\theta \tag{13.13}$$

Thus by the definition of b we obtain

$$\frac{a(e^{j\theta})}{|a(e^{j\theta})|(e^{j\theta})^n} = \frac{F(e^{j\theta})}{|F_1(e^{j\theta})|}, \qquad \forall\theta \tag{13.14}$$

i.e.,

$$a(e^{j\theta}) = \frac{|a(e^{j\theta})|}{|F_1(e^{j\theta})|}(e^{j\theta})^n F_1(e^{j\theta}), \qquad \forall\theta \tag{13.15}$$

Now set $F = (e^{j\theta})^{(n-1)} F_1(e^{j\theta})$ and $\lambda = \frac{|a(e^{j\theta})|}{|F(e^{j\theta})|}$. Note that since by hypothesis $n \geq 1$, we have that $F \in A$. This proves half of the Lemma. The other direction is obvious. ■

We now try to solve (13.12). The unknowns in equation (13.12) are the functions λ, F and f and the scalar γ. The difficulty with applying Newton's method to equation (13.12) is that, in a sense, it has more unknown functions (four) than constraints (two). This means that local solutions are not isolated, rather there is a whole continuum of them. Newton's method is well known for its bad behavior in such circumstances, because in the linearization (13.11) the differential T'_{x_k} will not be invertible. Consequently, rather than working with (13.12) directly, we modify it first to obtain a problem with only two unknowns. This is the key step in our method to finding solutions to (13.12).

We proceed with formal calculations for a moment. Begin by eliminating indeterminacy in (13.12) by assuming $\lambda^{-1} \in L^2_1$ is such that $\int_0^{2\pi} \lambda(e^{j\theta})^{-1} \frac{d\theta}{2\pi} = 1$. Note that $P^-[\ \Gamma(\cdot, f)\ ] = 0$ since $\Gamma(\cdot, f)$ is a constant function, and that $P_0^-[\ e^{j\theta} F\ ] = 0$ (provided $F \in H^2$). Thus we obtain the system

$$\begin{aligned} P^-[\ \Gamma(\cdot, f)\ ] &= 0 \\ P_0^-[\ \lambda^{-1}\ \tfrac{\partial \Gamma}{\partial z}(\cdot, f)\ ] &= 0 \end{aligned} \qquad (13.16)$$

in two unknowns, λ and f. Because of the way λ is normalized, one can write

$$\lambda^{-1} = 1 + \Re(e^{j\theta}\beta) \qquad (13.17)$$

where β is a scalar valued function which is analytic on the disk. Now, substituting (13.17) into (13.16) gives an operator equation of the form

$$T(f, \beta) = \mathbf{0} \qquad (13.18)$$

At this point the basic idea is clear: to find solutions to the optimality conditions (i) and (ii) of Theorem 13, it is enough to solve equation (13.18). Thus we state the following:

Critical Point Problem.

Find analytic functions f^*, β^* with $1 + 2\Re(e^{j\theta}\beta^*)$ positive on the circle such that (f^*, β^*) is a solution to equation (13.18).

We have found that Newton's method applied to the Critical Point Problem is extremely effective:

Newton's Method Given (f, β). Solve the equation

$$T(f, \beta) + T'_{(f,\beta)}[\epsilon, \eta] = \mathbf{0}$$

for ϵ, η and produce an update $(f + \epsilon, \beta + \eta)$.

A standard fact about Newton iteration is that when the linearizations $T'_{(f,\beta)}$ are uniformly invertible, the iterates have an excellent convergence property called second order convergence.

This is analysed in **[HMW]** where the linearizations are shown to be invertible and consequently our algorithm is second order convergent. It requires quite a bit of functional analysis to find the correct setting and estimates.

While it is well beyond the scope of our presentation here, we give the formula for T' in Chapter 16.

Chapter 14

References

The optimization problem of approximating in the norm of the supremum a given continuous function on the circle by functions in the disk algebra is a primitive version of the problem OPT considered in this part of the book. As such it was treated first by Nehari in 1949, who proved that the distance is the largest singular value of certain Hankel operator. Many workers made contributions, one specially worth mentioning is the team Adamjan-Arov-Krein, who developed an extensive theory in a series of papers [**AAK1-AAK3**].

The mathematical theory described in this part of the book started with Helton and Howe [**HH**]. Theorem 13 was stated first and proved in [**H6**]. Theoretical work and other results pertaining Theorem 13 can be found in [**HMer2**], [**M**], [**Hu**], [**HMar**].

Numerical work describing steepest descent type of algorithms for solving OPT is in [**HMer3**]. The algorithm described here is in [**HMW**].

[**A**] Ahlfors. Complex Analysis.

[**AAK1**] V. M. Adamjam, D. Z. Arov and M. G. Krein. Infinite Hankel matrices and generalized problems of caratheodory-Fejer and F. Riesz, Functional Anal. Appl. 2 (1968), pp. 1-18.

[**AAK2**] V. M. Adamjam, D. Z. Arov and M. G. Krein. Analytic properties of Schmidt paris for Hankel operator and the generalized Scur-Takagi problem. Math. USSR-Sb. 15 (1972), pp. 15-78.

[**AAK3**] V. M. Adamjam, D. Z. Arov and M. G. Krein. Infinite block Hankel matrices and related extension problems, Amer. Math. Soc. Translations 111 (1978), 133-156.

[H1] J. W. Helton. "Operator theory and broadband matching," Proceedings of Eleventh Annual Allerton Conference on Circuits and Systems Theory, 1976.

[H2] J. W. Helton. "A mathematical view of broadband matching," IEEE International Conference on Circuits and Systems Theory, New York, 1978.

[H3] J. W. Helton. "Broadbanding gain equalization directly from data," IEEE Trans. Circuits and Systems, CAS-28 (1981), No. 12, 1125–1137.

[H4] J. W. Helton. "An H-infinity approach to control," IEEE Conference on Decision Control, San Antonio, Texas, December 1983

[H5] J. W. Helton. "Worst case analysis in the frequency domain: The H-infinity approach to control," IEEE Trans. Auto. Cont., AC-30 (December 1985), No. 12, 1154–1170.

[H6] J. W. Helton. "Optimization over spaces of analytic functions and the Corona problem," J. Operator Theory, 15 (1986), No. 2, 359–375.

[H7] J. W. Helton: Optimal frequency domain design vs. an area of several complex variables, pleniary address. MTNS, 1989.

[HH] J. W. Helton and R. Howe. "A bang-bang theorem for optimization over spaces of analytic functions," J. Approx. Theory, 47 (1986), No. 2, 101–121.

[HMar] J. W. Helton and D. Marshall: "Frequency domain design and analytic selections," Indiana Univ. Math. Journal, 39 (Spring 1990), No. 1, 157–184.

[HMer1] J. W. Helton and O. Merino: "Optimal Analytic Disks," Collection: Several complex variables and complex geometry, Part 2, (Santa Cruz, CA, 1991) Proc. Sympos. Pure Math., 5 251–262.

[HMer2] J. W. Helton and O. Merino. Conditions for Optimality over $\mathcal{H}^\infty$. SIAM J. Cont. Opt., Vol. 31, No. 6, November 1993.

[HMer3] A Novel Approach to Accelerating Newton's Method for sup-norm optimization arising in H^{∞}-control. J. Optim. Theory Apps. 1993

[HMW] J. W. Helton, O. Merino and T. E. Walker. Algorithms for Optimizing over Analytic Functions. Indiana U. Math. J., Vol 42, No.3 (1993) .

[HV] J. W. Helton, and A. Vityaev. Preprint

[Hof] K. Hoffmann. Banach spaces of analytic functions. Dover

[Hui] S. Hui. Qualitative Properties of Solutions to H^{∞}-Optimization. Problems. J. of Functional Analysis **75** (1987) 323,348.

[M] O. Merino. Optimization over spaces of analytic functions. Thesis. University of California, San Diego. 1988.

[R] W. Rudin. Real and complex analysis.

Part IV

H^{∞} Theory: Vector Case

Chapter 15

Many Analytic Functions: Optimization for MIMO system design

The design methods presented in this book extend directly to multiple input, multiple output (MIMO) systems. Indeed one can reformulate design problems as optimization problems with interpolation constraints on matrix-valued analytic functions. Parameterizations of the type

$$T(s) = T_1(s) + T_0(s)F(s), \tag{15.1}$$

discussed in Chapter 8 for the scalar case, also arise but now the functions involved are matrix valued analytic functions. Just as we had in the scalar case, design problems correspond to OPT problems, but this time over N-tuples of analytic functions $(f_1, f_2, \ldots, f_n)$ instead of a single analytic function. These are the most challenging problems which arise in MIMO design, and fortunately much is known about them.

In this chapter, while we do not lay out MIMO control theory, we do give one the tools to solve the H^∞ optimization problem that arises from this subject. This tool is the computer program Anopt. It is written to optimize over analytic functions on the unit disk rather than the R.H.P. Anopt applies most directly to the design of discrete time systems. We warn the reader that the constrained optimization required to codify internal stability in the vector case requires additional programming, for example to produce automatically a formula analogous to (15.1).

We include in the appendix the Mathematica notebook AnoptTutorial.ma, which is an introductory manual for using the program Anopt.

15.1 The OPT problem

We denote with A_N the (real) vector space of all N-tuples $f = (f_1, \dots, f_N)$ where f_ℓ is in A. The space A_N is a complete normed space (or Banach space), equipped with the norm

$$||f||_\infty = \max_\theta ||F(e^{j\theta})||_{C^N},$$

where $||z||_{C^N}$ represents the euclidean norm on C^N. A performance function is a positive valued function Γ on $\partial D \times C^N$. We will use the notations $\Gamma(e^{j\theta}, z)$ and $\Gamma(e^{j\theta}, z_1, \dots, z_N)$ to denote the same object. The optimization problem OPT over the set A_N is

OPT Given a performance function $\Gamma(e^{j\theta}, z)$, find $\gamma^* \geq 0$ and f^* in A_N (if it exists) such that

$$\gamma^* = \inf_{f \in A_N} \sup_\theta \Gamma(e^{j\theta}, f(e^{j\theta})) = \sup_\theta \Gamma(e^{j\theta}, f^*(e^{j\theta})).$$

The function f^* in the statement OPT is called a *solution* or *minimizer*.

The theoretical study of OPT with most practical implications is the characterization of solutions. We already saw in the study of the scalar case how conditions for optimality lead quickly to practical tests for optimality, and to algorithms. The situation in the vector case is quite similar, and this will become apparent in the following sections.

15.2 Solutions only a keystroke away

The theory and algorithms of OPT are mature enough to permit the use of computers to obtain approximate solutions as well as diagnostics with few keystrokes. This might look like an exaggeration[1], so to convince the reader we present a computer solved example.

Consider the performance function

$$\begin{aligned} \Gamma(e^{j\theta}, \quad & z_1, z_2) = \\ & |\tfrac{1}{e^{j\theta}+0.5} - z_1|^2 + 2\mathrm{Im}(z_1 \; z_2) \qquad (15.2) \\ & +(5 + \mathrm{Re}(\tfrac{1}{2+e^{j\theta}}))|\tfrac{1}{e} + z_2|^2 \end{aligned}$$

The following is the input the user types in a Mathematica session in order to solve OPT.

---

[1] indeed it is an exageration in cases which cause numerical difficulties.

```
<<Anopt`;

g = Abs[1/(e + 0.5 ) - z[1] ]^2
    + ( 5 + Re[1/(2 + e)] ) Abs[ 1/e + z[2]]^2
    + 2 Im[z[1] z[2]]^2;

Anopt[g,0.000001];
```

In the session above the input is the performance "g" and the number 0.01, which is an error tolerance for stopping calculations. The following is output that is printed on the screen:

```
It :    Current Value    :  Step   : Optimality Tests  :  Error  : Sm. : Grid
   :      SupGamma       :         :   OT1       OT2   :   NED   :     :
-----------------------------------------------------------------------------
 0 : 1.             E+01 :   N/A   : 4.2E-01 : 1. E-02 :   N/A   : NON : 32
 1 : 6.9217870409218E+00 : 2.3E-01 : 2.3E-04 : 7.1E-04 : 5. E-02 : NON : 32
 2 : 6.9214867847569E+00 : 1.9E-03 : 2.4E-05 : 4.7E-05 : 7.1E-04 : NON : 32
 3 : 6.921495036711 E+00 : 0  E+00 : 2.6E-05 : 4.7E-05 : 7.1E-04 : 5A10: 32
 4 : 6.921446086706 E+00 : 1.3E-04 : 4. E-07 : 1.2E-06 : 1.7E-07 : NON : 64
 5 : 6.9214453945937E+00 : 3.4E-06 : 4. E-10 : 1.3E-08 : 2.6E-07 : NON : 64


Summary
-------
Supremum of gamma: SupGamma = 6.921445394593699E+00
Optimality Test  : OT1      = 4.0149612824E-10
Optimality Test  : OT2      = 1.3493172843E-08
Error diagnostic : NED      = 2.6 E-07
```

The final value of SupGamma is the calculated value of γ^* in the statement OPT. The numbers under the headings OT1 and OT2 are the result of optimality tests. At the solution, they should be equal to zero. The analytic function that is the calculated solution is also an output, but is not printed. The reader interested in more details on the output and on computer solutions may read the notebook AnoptTutorial.ma, which is an appendix in this book. Also OT1 and OT2 are based on the theory developed in the next three sections.

15.3 Necessary conditions for solutions

As in the scalar case (Theorem 13), diagnostics and computer algorithms in the vector case are based on a theorem characterizing solutions to OPT.

We now present a result that gives two necessary conditions for a solution. These two conditions already give enough information to derive an algorithm (see [**HMW**]). Additional conditions to insure sufficiency will be presented in Section 17.1.

So far we have mentioned "solutions f in A_N to OPT" that are global solutions. However in the study of conditions for optimality one must consider *local* solutions. This is needed because when $N > 1$, there is no uniqueness of solutions.

A *local solution* in A_N is a solution to OPT in some neighborhood $V \subset A_N$ of f. It is a *strict local solution* if it is unique in V. A function $h \in A_N$ is a *descent direction* relative to $f \in A_N$ if for all $t > 0$ small enough,

$$\sup_\theta \Gamma(e^{j\theta}, f(e^{j\theta}) + th(e^{j\theta})) < \sup_\theta \Gamma(e^{j\theta}, f(e^{j\theta}))$$

The function f is a *directional solution* or *directional minimizer* if there are no descent directions relative to f. Of course, every strict local solution is a directional solution.

Recall that Γ is a function of z where $z = (z_1, \ldots, z_N)$, so we can take partial derivatives such as $\frac{\partial \Gamma}{\partial z_1}(e^{j\theta}, z_1, \ldots, z_N)$ and $\frac{\partial \Gamma}{\partial z_2}(e^{j\theta}, z_1, \ldots, z_N)$.

Theorem 20 *Let Γ be of class C^3 and $f^* = (f_1, \ldots, f_N)$ in A_N be such that*

$$a(e^{j\theta}) := \left(\frac{\partial \Gamma}{\partial z_1}(e^{j\theta}, f^*(e^{j\theta})), \ldots, \frac{\partial \Gamma}{\partial z_N}(e^{j\theta}, f^*(e^{j\theta})) \right) \neq (0, \ldots, 0), \qquad \forall \theta$$

If the function f^* is a directional solution to OPT *then*

I $\Gamma(e^{j\theta}, f^*(e^{j\theta}))$ *is a constant function of $e^{j\theta}$.*

II *There exist $F = (F_1, \ldots, F_N)$ with F_ℓ analytic on the unit disk D and integrable boundary values $F_\ell(e^{j\theta})$, and there exists a positive valued function $\lambda(e^{j\theta})$ such that*

$$\begin{array}{ccc} \frac{\partial \Gamma}{\partial z_1}(e^{j\theta}, f^*(e^{j\theta})) & = & \lambda(e^{j\theta}) e^{j\theta} F_1(e^{j\theta}) \\ \vdots & \vdots & \vdots \\ \frac{\partial \Gamma}{\partial z_N}(e^{j\theta}, f^*(e^{j\theta})) & = & \lambda(e^{j\theta}) e^{j\theta} F_N(e^{j\theta}) \end{array}$$

15.4 An Example

Consider the performance function

$$\begin{aligned}\Gamma(e^{j\theta}, z_1, z_2) = |z_1|^2 \quad & +|z_2|^2 + |100 + e^{j\theta} z_1 + .1(z_1 + z_2 + z_1 z_2)|^2 \\ & +|100 + e^{j\theta} z_2 + .1(z_1 + z_2 + z_1 z_2)|^2\end{aligned} \tag{15.3}$$

This is a good example because, as we shall see, *there are two strict local minimizers f_1 and f_2, and there exists also a function f_0 that is a sort of " saddle point", i.e., f_0 satisfies conditions (I) and (II) but it is not a minimizer.*

In order to decide whether a candidate is or is not a solution we need sufficient conditions for optimality. Basically the sufficient conditions consist of (I), (II), and an additional condition that will be introduced later. For now we will only verify (I) and (II), leaving the rest to a later section.

The three functions are the following:

$$\begin{aligned} f_0(e^{j\theta}) &= (0,0) \\ f_1(e^{j\theta}) &= (30,-30) \\ f_2(e^{j\theta}) &= (-30,30) \end{aligned} \tag{15.4}$$

The calculations for the function f_2 are almost identical to those for f_1, so we present only those corresponding to f_1. We have,

$$\Gamma(e^{j\theta}, f_1(e^{j\theta}))) = \Gamma(e^{j\theta}, 30, -30) = 3800 \qquad \forall e^{j\theta}$$

so f_1^* satisfies (I). To check (II) we begin by calculating partial derivatives:

$$\begin{aligned} \tfrac{\partial\Gamma}{\partial z_1}(e^{i\theta}, z_1, z_2) \quad = & \\ & \overline{(100 + z_1 e^{i\theta} + 0.1(z_1 z_2 + z_1 + z_2)} \cdot (e^{i\theta} + 0.1 + 0.1 z_2) \\ & + \overline{(100 + z_2 e^{i\theta} + 0.1(z_1 z_2 + z_1 + z_2)} \cdot (0.1 + 0.1 z_2) + \overline{z_1} \\ \tfrac{\partial\Gamma}{\partial z_2}(e^{i\theta}, z_1, z_2) \quad = & \\ & \overline{(100 + z_1 0.1(z_1 z_2 + z_1 + z_2))}(0.1 + 0.1 z_1) \\ & + \overline{(100 + z_2 e^{i\theta} + 0.1(z_1 z_2 + z_1 + z_2))}(e^{i\theta} + 0.1 + 0.1 z_1) + \overline{z_2} \end{aligned} \tag{15.5}$$

Therefore,

$$\begin{aligned} \tfrac{\partial\Gamma}{\partial z_1}(e^{i\theta}, 30, -30) &= 10 e^{j\theta} + 2 \\ \tfrac{\partial\Gamma}{\partial z_2}(e^{i\theta}, 30, -30) &= 10 e^{j\theta} + 2 \end{aligned} \tag{15.6}$$

Note that

$$10e^{j\theta}+2 \;=\; e^{j\theta}\,|1+2e^{j\theta}|^2\,\frac{1}{10+2e^{j\theta}} \qquad (15.7)$$

and this implies that condition (II) is satisfied by f_1 with $\lambda(e^{j\theta}) = |1+2e^{j\theta}|^2$ and $F_\ell(e^{j\theta}) = 1/(10+2e^{j\theta})$, for $\ell = 1,2$.

Now calculations with f_0 are next. Condition I holds since

$$\Gamma(e^{i\theta},0,0) = 20,000 \quad (\forall\, e^{i\theta})\,.$$

To check II, evaluate the partial derivatives of Γ to obtain

$$\left(\frac{\partial\Gamma}{\partial z_1}(e^{i\theta},0,0),\, \frac{\partial\Gamma}{\partial z_2}(e^{i\theta},0,0)\right) = (100\,e^{i\theta}+20, 100\,e^{i\theta}+20)\,.$$

Note that

$$100\,e^{i\theta}+20 = 100\,|e^{i\theta}+0.2|^2\,\frac{e^{i\theta}}{1+0.2e^{i\theta}}$$

so II holds with $\lambda(e^{j\theta}) = 100\,|e^{i\theta}+0.2|^2$ and $F_\ell(e^{j\theta}) = e^{j\theta}/(1+0.2e^{j\theta})$, $\ell = 1,2$.

15.5 Computer diagnostics

Diagnostics and computer algorithms for the vector case follow immediately from the previous two sections and are a direct generalization of those for the scalar case found in Section 13.

Condition I of Theorem 20 can be easily implemented on a computer, by calculating

$$OT_1(f) = \frac{\sup_\theta\ \Gamma(e^{j\theta},f(e^{j\theta})) - \inf_\theta\ \Gamma(e^{j\theta},f(e^{j\theta}))}{\sup_\theta\ \Gamma(e^{j\theta},f(e^{j\theta}))}$$

To test condition II directly one also needs functions λ and F. If candidates for these functions are not already available, one would have to test every possibility in λ and F, an impossible task. Thus it is desirable to have a test of (II) that does not involve λ and F. The following conjecture is crucial in this respect:

Conjecture. For generic Γ condition II of Theorem 2 is equivalent to:

> Let $a_j = \frac{\partial\Gamma}{\partial z_j}(\cdot, f^*)$. Fix j_0. For each $j = 1,\ldots,N$, the function $\frac{a_j}{a_{j_0}}$ is analytic in the disk.

It is easy to check on the computer when this should apply. Thus the conjecture above allows one to check (for most performance functions) condition (II) of Theorem 20 by testing if each a_j/a_0 is analytic, provided $N > 1$. When $N = 1$ this test does not apply as stated, however it does reduce to the winding number test of Section 13.

To implement condition II, consider the Fourier expansion of the functions

$$\frac{a_\ell}{a_{\ell_0}} = \sum_{n=-\infty}^{\infty} c_{\ell,n} e^{jn\theta}$$

and form

$$OT_2(f) = (\sum_{\ell=1}^{N} \sum_{n<0} |c_{\ell,n}|^2)^{1/2}$$

which is calculated on the computer with the Fast Fourier Transform. At optimum both OT_1 and OT_2 should equal zero to nearly machine precision. The program *Anopt* prints these values out at each iteration to monitor how the approximate solutions are progressing. If f is near the solution we must have $OT_1(f) \approx 0$ and $OT_2(f) \approx 0$.

To illustrate the practical use of the diagnostics we consider the performance function

$$\begin{aligned} \Gamma(e^{j\theta}, z_1, z_2) = \quad & |z_1|^2 + |z_2|^2 + |100 + e^{j\theta} z_1 + .1(z_1 + z_2 + z_1 z_2)|^2 \\ & + |100 + e^{j\theta} z_2 + .1(z_1 + z_2 + z_1 z_2)|^2 \end{aligned} \tag{15.8}$$

that was studied in Section 15.4. We already know that $f_1(e^{j\theta}) = (30, -30)$ and are directional solutions, and that $f_0(e^{j\theta}) = (0, 0)$ gives a saddle point (I and II are satisfied, but f_0 is not a solution).

Implementing the algorithm *Newton Iteration* (see sections 13.3 and 16) yields computer output, some of which we present below. The output shown in Tables 1 and 2 (except for the last column) was produced by the program Anopt written in Mathematica.

Initializing the iteration at $f_0(e^{j\theta}) = (29.6 + .1e^{j\theta}, -30.4 - .0001e^{j\theta} + .001(e^{j\theta})^2)$ produces the results indicated in Table 1 below. Since in this example we know what the solution is, we include in Table 1 a column that shows the true error $||f^k - f^*||_\infty$.

Table 1

```
            Value           Optimality Tests      True error
Iter      Sup Gamma           OT1       OT2         ||fk-f*||

 0   3.8165268771511E+03   8.3E-03   1.6E+00       5.7E-01
 1   3.8009177903975E+03   4.9E-04   3.9E-01       5.3E-02
 2   3.8000812563274E+03   4. E-05   1.5E-02       2.7E-03
 3   3.8000001798719E+03   6.3E-08   2. E-05       4.4E-06
 4   3.8000000000006E+03   2.4E-13   3.6E-10       1.5E-11
```

In the next example the Newton iteration exhibits second order convergence to the constant function $f^*(e^{j\theta}) = (0, 0)$, which is not a directional minimizer. The iteration was initialized with $f(e^{j\theta}) = (0.2 + 0.1e^{j\theta}, 1.4 + .1e^{j\theta} + 0.01(e^{j\theta})^2)$.

Table 2

```
            Value           Optimality Tests      True error
Iter      Sup Gamma           OT1       OT2         ||fk-f*||

 0   2.0428618081601E+04   3.2E-02   2.9E-01       1.4E+00
 1   2.0057600281609E+04   5.7E-03   9.3E-02       1.5E+00
 2   2.0097292539204E+04   1.2E-02   3.6E-02       7. E-01
 3   2.000858365994 E+04   6.8E-04   3.4E-03       3.1E-02
 4   2.0000014412272E+04   1.4E-06   6.6E-06       4.8E-05
 5   2.0000000000109E+04   9.7E-12   7.3E-11       7.6E-10
```

In these tests we have illustrated that, even in this rather elaborate function space setting, Newton‘s method behaves exactly as it would in traditional R^n optimization. Namely, it converges to a place where the "gradient" vanishes[2], be it a maximum, a minimum, or a saddle point.

15.6 Algorithms for OPT

The algorithms for OPT discussed in this book are instances of the following general scheme for solving nonlinear problems iteratively:

General Iteration Scheme

1. Test f for optimality. If optimal, then stop.
2. Solve a subproblem to obtain an update direction h.
3. Perform a search on $f + th$ with $t \geq 0$ to find a $t = t_0$ that satisfies a preset criterion.

---

[2] In our setting, this is conditions I and II of Theorem 20 are satisfied.

4. Set $f = f + t_0 h$, and go to step 1.

The tests for optimality OT_1 and OT_2 we have developed are used in Step 1. Another stopping criterion one may want to include in particular implementations is "lack of sufficient progress" (however you want to define this). A typical Step 2 consists of replacing the performance function with a "model", a new performance which is based on the original function but easier to deal with. Step 3 is called *line search*. One obvious example of the linesearch criterion is that t minimizes $\sup_\theta \Gamma(e^{j\theta}, f(e^{j\theta}) + th(e^{j\theta}))$.

The two algorithms discussed later in this book (see Chapter 16) are *Newton Iteration* and *Disk Iteration*. Both are based on models of the performance obtained by taking some terms from the Taylor expansion of $\Gamma(e^{j\theta}, z)$ about $z =$ the current guess f.

This section addresses what we have found to be the most effective means of evaluating the performance of a computer algorithm. The objective is to have a theoretical test, which can be done with pencil and paper at the same time that you are developing an algorithm. This is tremendously helpful in discovering computer algorithms. the reason is that the algorithm itself is presented analytically and having an analytic means of evaluation allows one to see how to make changes which improve performance. What we recommend as the leading indicator of success of an algorithm is actually very conventional. It is the *order of convergence* .

A sequence of vectors x_n in a vector space with norm $\| \ \|$ that converges to x_* is said to be p-convergent if there exists $C > 0$ such that

$$\text{(EST)} \qquad \|x_n - x_*\| < C\|x_n - x_*|^p$$

The larger p the faster the sequence converges. An algorithm which generates sequences will be called *order p convergent* provided the sequences it generates are at least *order p convergent.* Here C is a number whose size influences the error estimate (EST) less than p. When $p = 1$ we need $C < 1$ to guarantee improvement from one iteration to the next.

A summary of the performance of algorithms presented in this book is:

Method	*Case*	*Convergence*
Disk Iteration	$N = 1$	order 2
	$N > 1$	order 1
Newton Iteration	all N	order 2

Some technical points:

- Convergence of the Disk Iteration algorithms has not been established. However one can show that if it converges, then the limit function f^* must satisfy the optimality conditions I and II. *We conjecture that if Disk Iteration with a line search converges to f^*, then f^* is a local solution of OPT.* The difficulty with establishing convergence is that in the function spaces over which the optimization is performed, bounded sets are not necessarily compact. In such spaces the distance between elements of a sequence may decrease to 0 and the sequence may still not converge. Our experience indicates that formal order of convergence is easier to analyze and of much more practical importance in predicting what you see on a computer than actual convergence itself.
- We have not specified which norm we are using in the table. The details are rather complicated and do not contribute to our exposition.
- Newton iteration may converge to an f^* which is not a local minimum, but a sort of critical point. See **[HMer1-2]**.

15.7 References

The main theorem in this section is from **[HMer2]**. More results can be found in **[H3]** and in **[HMar]**. See **[H1]** for connections with other areas of Mathematics. Several Complex Variables connections are discussed in **[HMer1]**, see also **[HH]**, **[HMar]** and **[HV]**.

[H1] J. W. Helton. Operator Theory, Matrices and Electrical Engineering. NSF. 1987

[H2] J. W. Helton. "Optimization over spaces of analytic functions and the Corona problem," J. Operator Theory, 15 (1986), No. 2, 359–375.

[H3] J. W. Helton: Optimal frequency domain design vs. an area of several complex variables, plenary address. MTNS, 1989.

[HH] J. W. Helton and R. Howe. "A bang-bang theorem for optimization over spaces of analytic functions," J. Approx. Theory, 47 (1986), No. 2, 101–121.

[HMar] J. W. Helton and D. Marshall: "Frequency domain design and analytic selections," Indiana Univ. Math. Journal, 39 (Spring 1990), No. 1, 157–184.

[HMer1] J. W. Helton and O. Merino: "Optimal Analytic Disks," Collection: Several complex variables and complex geometry, Part 2, (Santa Cruz, CA, 1991) Proc. Sympos. Pure Math., 5 251–262.

[HMer2] J. W. Helton and O. Merino. Conditions for Optimality over $\mathcal{H}^\infty$. SIAM J. Cont. Opt., Vol. 31, No. 6, November 1993.

[HV] J. W. Helton, and A. Vityaev. Preprint

Chapter 16

More Numerical Algorithms

This chapter is one of the more theoretical parts of the book which is ironic in that it derives the formulas which are used in our computer algorithms. The authors are a bit iconoclastic in that they see developing computer algorithms of this type as almost an exercise in pure mathematics. Indeed the next appendix describes pure mathematics implications of this and previous sections. A design engineer can skip this section. An engineer who develops optimization methods might find it interesting.

In this chapter we discuss in some detail two algorithms for solving OPT , Disk Iteration and Newton Iteration. The discussion begins with the statement of classical results on distance problems in Section 16.2. Disk Iteration and Newton algorithms are presented in Sections 16.3 and 16.4. The chapter ends with Section 16.5 which treats the connection between the classical results and the characterization of solutions to OPT.

16.1 Notation

We define the spaces L^p_N, H^p_N, whose elements are N-tuples of functions $(f_1, \ldots, f_N)$ and whose norms are defined in terms of the euclidean norm on C^N, as in

$$\begin{aligned} ||f||_{L^p_N} &= \left(\int_0^{2\pi} ||(f_1(e^{j\theta}), \ldots, f_N(e^{j\theta}))||^p_{C^N} \frac{d\theta}{2\pi}\right)^{\frac{1}{p}} \qquad (1 \le p < \infty) \\ ||f||_{L^\infty_N} &= \sup_\theta ||(f_1(e^{j\theta}), \ldots, f_N(e^{j\theta}))||_{C^N} \end{aligned} \tag{16.1}$$

If $f \in L^p_N$ for some $1 \leq p \leq \infty$, then the Fourier coefficients $\hat{f}_\ell$ are vectors in C^N.

The orthogonal complement in L^2_N of a subspace $\mathcal{S}$ is denoted $\mathcal{S}^\perp$. If k is an $M \times N$-matrix valued function on the unit circle, the Hankel operator with symbol k is the operator $\mathcal{H}_k : H^2_M \to H^2_{N,0}$ with action $\alpha \to P_{H^{2\perp}_{N,0}}[k\alpha]$. The Toeplitz operator with symbol k is the operator $\mathcal{T}_k : H^2_M \to H^2_N$ with action $\alpha \to P_{H^2_N}[k\alpha]$.

16.2 Nehari's problem: perspective on OPT from a classical viewpoint

The math problems treated and used in this book long preceded engineering application. Some go back to the beginning of this century. A variant which is close to the spirit of this book is the classical Nehari problem which dates from the 1950's.

Nehari Problem. Let k be a function in L^∞_N, find its distance to H^∞_N.

Here the distance of k to H^∞_N means

$$\text{dist}(k, H^\infty_N) := \inf_{f \in H^\infty_N} ||k - f||_{L^\infty_N} \tag{16.2}$$

and a *best H^∞ approximation to k* is a function $f_0 \in H^\infty_N$ that realizes the supremum in (16.2). Standard theory shows that f_0 exists. Moreover, if $k \in \mathcal{C} + H^\infty_N$ then the best approximation is unique (here $\mathcal{C}$ is the set of continuous functions on the unit circle). We refer to the problem of calculating $\text{dist}(k, H^\infty_N)$ and a minimizer $f_0 \in H^\infty_N$ as *Nehari's problem.* Nehari's problem is also the simplest nontrivial OPT problem with the performance

$$\Gamma(e^{j\theta}, z) = ||k(e^{j\theta}) - z||_{C^N} \tag{16.3}$$

Let $\mathcal{H}_k : H^2_1 \to {H^2_N}^\perp$ be the operator with action $\alpha \to P_{H^{2\perp}_N}[k\alpha]$. This type of operator is called a *Hankel* operator and k is called its symbol. A well known result is:

Theorem 21 *If k is in $\mathcal{L}^\infty_N$, then*

$$|| \, \mathcal{H}_k \, ||_{\mathcal{L}} \;=\; \text{dist}(k, H^\infty_N) \tag{16.4}$$

and there exists a best approximant $f_0 \in \mathcal{H}^\infty_N$. If α_0 is a function at which $\mathcal{H}_k$ attains its norm, then

$$k - f_0 = \frac{1}{\alpha_0} \mathcal{H}_k[\alpha_0] \tag{16.5}$$

If the symbol k is in $\mathcal{H}^\infty + \mathcal{C}$, then f_0 is unique. Let $\mathcal{H}_k^*$ be the adjoint of $\mathcal{H}_k$, and set $\tau = \mathrm{dist}(k, H_N^\infty)$; then

$$\mathcal{H}_k^*[\, \mathcal{H}_k[\alpha_0]\,] = \tau^2 \alpha_0 \tag{16.6}$$

Equation (16.6) says that α_0 is a function associated with τ, the largest singular value of the compact operator $\mathcal{H}_k$ (i.e., the norm of $\mathcal{H}_k$). We will refer to equation (16.6) as the *Nehari - Commutant Lifting formula.* For $N > 1$ this follows from commutant lifting arguments **[A]**, **[S]**, **[NF]**, **[AAK1–3]**.

16.3 Disk Iteration algorithms

The solution of Nehari's problem leads directly to numerical methods for solving OPT problems where the sublevel sets are disks ("disk problems"). To solve more general OPT problems we can successively approximate OPT by disk problems and solve them. We call this Disk Iteration. These are described together with experimental findings in **[HMer1]**, and here we only sketch the basic idea.

16.3.1 The Power Method for solving Nehari's Problem

A common way to solve Nehari's problem is to solve (16.6) for α_0 as an eigenvalue problem, and then solve for f_0 in (16.5). The power method is a natural choice since we are seeking the largest eigenvalue of the operator $\mathcal{H}_k^*\mathcal{H}_k$, and the corresponding eigenvector. It is a well known fact that, provided the multiplicity of this eigenvalue is equal to 1, the power method converges linearly to the solution.

The Power Method. Given k and a guess α a solution to (16.6), update α to α_1 with

P1. $\alpha_+ := P_{H^2_N}[\, \overline{k}^t P_{H^{2\perp}_N}[\, k\alpha\,]\,]$.

P2. $\alpha_1 := \frac{1}{\| \alpha_+ \|_{L^2_1}} \alpha_+$.

Suppose that the unit circle is discretized to n (a power of 2) equally spaced points, so that one can calculate projections in $2n\log_2(n)$ operations by using a Fast Fourier Transform algorithm (we do not count additions). Then step P1 takes $N(4n\log_2(n) + 2n)$ operations, and step P2 takes Nn

multiplications, 1 square root and Nn divisions. Thus the power method for solving Nehari's problem is of order $Nn\log_2(n)$. Also, note that memory requirements are low since no matrices have to be stored at any moment.

For comparison we mention that Newton Iteration requires the inversion of a matrix M that arises from a discretization of the operator $T'_{f,\beta}$ introduced in Sections 13.3 and 16.4. The matrix M has a number of entries that scale with Nn^2, but it has special structure that can be exploited to reduce the operation count in the inversion (with algorithms of the type found in **[SLK]**), possibly reducing the order of the inversion to Nn^2. Also there exist algorithms for inverting a matrix by iteration, which can be effective in this setting.

The power method is clearly efficient in many respects, and it is superior to the Newton Iteration in memory requirements and operation count. Obviously it is also better when it comes to time actually used to produce an update in step 2 of the iteration scheme. However, there is a crucial limitation of the power method: it is directly applicable only to very simple problems OPT, namely Nehari problems.

The properties of the power method make it an interesting tool for solving OPT problems through approximating Γ by Taylor expansions. This Taylor expansion is used to obtain a Nehari problem, or at least a problem close to it. This has been attempted before in **[HMer1]**. In Subsection 16.3.2 below the method of *Disk iteration*, see **[HMer1]**, is described.

16.3.2 The Algorithm

We now introduce a Disk Iteration algorithm.

Disk Iteration. Given a current guess at the solution $f \in A_N$, the Step 2 of the Scheme is:

- Find $h = (h_1, h_2, \ldots, h_N) \in A_N$ that minimizes

$$\sup_{e^{j\theta}} \left\{ \Gamma(\cdot, f) + 2 \text{ Re } \left(\frac{\partial \Gamma}{\partial z}(\cdot, f)^t h \right) + \sum_{k=1}^{N} \frac{\partial^2 \Gamma}{\partial z_k \partial \overline{z_k}}(\cdot, f) \; |h_k|^2 \right\} \tag{16.7}$$

A first glance at the algorithms reveals that the Taylor expansion of Γ about f, the current guess at the solution, is not necessarily of the form of a Nehari problem. However, it can be solved by iterating Nehari problems **[HMer1]**. Thus an implementation of the Disk iteration algorithm that uses the power method requires an inner loop to find a solution to Step 2 [1].

Another feature of (16.7) is that the Taylor expansion it contains does not capture all second order information from most functions Γ. One expects that this would imply first order (bad) convergence for the algorithm. This is indeed true in the vector valued case $N > 1$, but it is surprising that (in practice) second order (good) convergence occurs in the scalar case $N = 1$ (see **[HMer1]**).

Thus Disk iteration gives second order convervence in practice when $N = 1$, and no better than first order when $N > 1$.

16.4 More on the Newton iteration algorithm

This section describes Newton's method of solving OPT when $N > 1$. We follow closely the reference **[HMW]**. Recall that in the scalar case (Section 13.3) the key was a set of Critical Point equations. The vector case (C^N-valued analytic functions) of the Critical Point equations is:

$$\begin{aligned} P_{H^{2\perp}_{N,0}}\left[(1+2\ \mathrm{Re}\ (\chi\beta))\ \tfrac{\partial\Gamma}{\partial z}(\cdot,f)\right] &= 0 \\ P_{H^{2\perp}}\left[\ \Gamma(\cdot,f)\ \right] &= 0 \end{aligned} \tag{16.8}$$

where $\lambda^{-1} = 1 +\ \mathrm{Re}\ (\chi\beta)$ and where β is a scalar valued function which is analytic on the disk. Thus we get the nonlinear equation

$$T\begin{pmatrix} f \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{16.9}$$

where in equation (16.9),

$$T\begin{pmatrix} f \\ \\ \beta \end{pmatrix} = \begin{pmatrix} P_{H^{2\perp}_{N,0}}\left[(1+2\ \mathrm{Re}\ (\chi\beta))\ \tfrac{\partial\Gamma}{\partial z}(\cdot,f)\right] \\ \\ P_{H^{2\perp}}\left[\ \Gamma(\cdot,f)\ \right] \end{pmatrix} \tag{16.10}$$

Thus our objective is to solve the Critical Point Problem (16.9) with T given by (16.10). Here χ is the function on the unit circle defined by $\chi(e^{j\theta}) = e^{j\theta}$.

Now we turn to the solution of operator equations (e.g. equation (16.9)) with the iterative method of Newton. In terms of T' we have the Newton's

---

[1] We note that there are many reasonable modifications of the expression (16.7) that preserve intact terms of order up to 1 in h. In fact in **[HMer1]** one was found which eliminates the inner loop at the expense of *all* the second order information.

step for updating a given pair (f, β):

$$\begin{pmatrix} \tilde{f} \\ \tilde{\beta} \end{pmatrix} = \begin{pmatrix} f \\ \beta \end{pmatrix} - T'^{-1}_{f,\beta} T \begin{pmatrix} f \\ \beta \end{pmatrix} \tag{16.11}$$

We refer to (16.11) as *Newton iteration*, and to the repeated application of (16.11) as the *Newton algorithm.*

We now derive the formula for the Jacobian $T'_{f,\beta}$ of T, which we obtain by linearizing (16.10).

$$\begin{array}{rcl} \Gamma(\cdot, f+\epsilon) & = & g + 2\ \mathrm{Re}\ (a^t\epsilon) + \overline{\epsilon}^t A\epsilon + \ \mathrm{Re}\ (\epsilon^t B\epsilon) + \cdots \\ \frac{\partial\Gamma}{\partial z}(\cdot, f+\epsilon) & = & a + A\overline{\epsilon} + B\epsilon + \cdots \end{array} \tag{16.12}$$

where

$$g = \Gamma(\cdot, f), \quad a = \tfrac{\partial\Gamma}{\partial z}(\cdot, f), \quad A = \tfrac{\partial^2\Gamma}{\partial\overline{z}\partial z}(\cdot, f), \quad \mathit{and} \quad B = \tfrac{\partial^2\Gamma}{\partial z^2}(\cdot, f). \tag{16.13}$$

Consider the expression

$$T\begin{pmatrix} f+\epsilon \\ \beta+\delta \end{pmatrix} = \begin{pmatrix} P_{H^{2\perp}_{N,0}}\left[\ (1 + 2\ \mathrm{Re}\ (\chi(\beta+\delta)))\ \frac{\partial\Gamma}{\partial z}(\cdot, f+\epsilon)\ \right] \\ P_{H^{2\perp}_{1}}\left[\ \Gamma(\cdot, f+\epsilon)\ \right] \end{pmatrix} \tag{16.14}$$

Substituting (16.12) into (16.14) and dropping all terms which are not linear or constant in δ or ϵ yields

$$T\begin{pmatrix} f \\ \beta \end{pmatrix} + T'_{f,\beta}\begin{pmatrix} \epsilon \\ \delta \end{pmatrix} =$$

$$\begin{pmatrix} P_{H^{2\perp}_{N,0}}\left[\ (1 + 2\ \mathrm{Re}\ (\chi(\beta+\delta)))\,(a + A\overline{\epsilon} + B\epsilon) + 2\ \mathrm{Re}\ (\chi\delta)a\ \right] \\ P_{H^{2\perp}_{1}}\left[\ g + 2\ \mathrm{Re}\ a^t\epsilon\ \right] \end{pmatrix} \tag{16.15}$$

Thus we have that the Jacobian of T is given by

$$T'_{f,\beta}\begin{pmatrix} \epsilon \\ \delta \end{pmatrix} = \begin{pmatrix} P_{H^{2\perp}_{N,0}}\left[\ (1 + 2\ \mathrm{Re}\ (\chi\beta))\,(A\overline{\epsilon} + B\epsilon) + 2a\ \mathrm{Re}\ (\chi\delta)\ \right] \\ P_{H^{2\perp}_{1}}\left[\ 2\ \mathrm{Re}\ a^t\epsilon\ \right] \end{pmatrix} \tag{16.16}$$

An elegant way to write T' appears in the following proposition:

Proposition 4 *The operator $T'_{f,\beta}$ in (16.16) can be written in the form*

$$T'_{f,\beta} = \begin{pmatrix} I_N & 0 \\ 0 & \overline{\chi} \end{pmatrix} \mathcal{C}\mathcal{T}_{M_1} + \begin{pmatrix} \chi I_N & 0 \\ 0 & 1 \end{pmatrix} \mathcal{H}_{M_2} \tag{16.17}$$

where $\mathcal{T}_{M_1}$ *is a Toeplitz operator with symbol* $M_1 = \begin{pmatrix} wA & \chi\overline{a} \\ \overline{\chi}a^t & 0 \end{pmatrix}$, $\mathcal{H}_{M_2}$ *is the Hankel operator with symbol* $M_2 = \begin{pmatrix} \overline{\chi}wB & a \\ a^t & 0 \end{pmatrix}$, $\mathcal{C}$ *is the operator* $f \rightarrow \overline{f}$, I_N *is the* $N \times N$ *identity matrix, and* $w = 1 + 2 \text{ Re } \chi\beta$.

A proof of Proposition 4 follows from the relations $P_{H^{2\perp}_{N+1,0}}[\overline{M\overline{\eta}}] = \mathcal{C}P_{H^2_{N+1}}[M\eta]$ and $\overline{\chi}P_{H^{2\perp}_{N+1,0}}[\chi M\eta] = P_{H^{2\perp}_{N+1}}[M\eta]$ applied to (16.16).

One can see in representation (16.17) that $T'_{f,\beta}$ has the form of a "Conjugate of Toeplitz plus Hankel" operator (except for a small change in the range, which turns out to be unimportant for our analysis). One consequence of this as pointed out to us by Ali Sayed is that $T'{}_{f,\beta}$ can be inverted numerically with fast algorithms similar to those described in **[SLK]**. This enhances the practical appeal of Newton's method applied to the operator T.

An important issue is whether or not $T'_{f,\beta}$ is invertible for almost all (f, β). This is required for Newton's method to work well. A technical point is the space on which $T'_{f,\beta}$ is defined. This certainly influences invertibility. For Newton iteration it is appropiate to define T is defined not on all of an H^2 type space but on a subspace of it consisiting of smooth functions. This causes mathematical problems in analyzing the behavior of T but not in computer implementation. See Theorem 6 in **[HMW]** for details.

16.5 Derivation of the Nehari-Commutant Lifting and other formulas

In this subsection we derive relations (16.4), (16.5), and (16.6) that appear in Theorem 21 directly from the optimality conditions in Theorem 20.

Recall that the Nehari problem is of the type OPT with performance as in (16.3). To avoid technical issues we assume that $k \notin H^\infty_N$ and that k is continuous, hence there is a unique best H^∞ approximation f_0 to k. We also assume that $k - f_0$ is bounded away from 0. Thus Γ and f_0 satisfy the hypotheses of Theorem 20.

Set $\tau = \text{dist}(k, H^\infty_N)$. The optimality conditions I and II of Theorem 20 in this case say that there exists λ measurable and positive on T, and a function $F \in H^1_N$ such that

$$||k(e^{j\theta}) - f_0(e^{j\theta})||^2_{C^N} = \tau^2 \tag{16.18}$$

$$\frac{1}{\lambda}\overline{(k - f_0)} = \chi F \tag{16.19}$$

Let $\alpha_0 \in H^2_1$ be the outer spectral factor of $1/\lambda$. From (16.19) we have that

$$k - f_0 = \frac{1}{\alpha_0} \frac{\overline{\chi F}}{\overline{\alpha_0}} = \frac{1}{\alpha_0} P_{H^{2\perp}_N} [\frac{\overline{\chi F}}{\overline{\alpha_0}}] \tag{16.20}$$

Combining (16.19) and (16.20) one obtains

$$k - f_0 = \frac{1}{\alpha_0} P_{H^{2\perp}_N} [(k - f_0)\alpha_0] = \frac{1}{\alpha_0} P_{H^{2\perp}_N} [k\alpha_0] \tag{16.21}$$

This proves the formula (16.5).

We now check the identity (16.4). We trivially have that

$$||\mathcal{H}_k||^2 = ||\mathcal{H}_{k-f_0}||^2 \ \leq \ ||k - f_0||^2_{L^\infty_N} = \tau^2 \tag{16.22}$$

To prove the reverse inequality, note that by (16.18) the following relation holds for almost all θ:

$$\tau^2 = \frac{|\alpha_0(e^{j\theta})|^2 ||k(e^{j\theta}) - f_0(e^{j\theta})||^2_{C^N}}{|\alpha_0(e^{j\theta})|^2} = \frac{||\alpha_0(e^{j\theta})(k(e^{j\theta}) - f_0(e^{j\theta}))||^2_{C^N}}{|\alpha_0(e^{j\theta})|^2} \tag{16.23}$$

Combine (16.5) and (16.23) to obtain

$$\tau^2 = \frac{||\alpha_0(k - f_0)||^2_{L^2_N}}{||\alpha_0||^2_{L^2_N}} = \frac{||P_{H^{2\perp}_N}[\alpha_0(k - f_0)]||^2_{L^2_N}}{||\alpha_0||^2_{L^2_N}} \leq \ ||\mathcal{H}_k||^2_{\mathcal{L}} \tag{16.24}$$

and this finishes the proof of the equality (16.4).

At this point we are in position of deriving the Nehari - Commutant Lifting formula (16.6). Rewrite (16.19) as

$$\overline{\alpha_0}(\overline{k - f_0}) = (\chi/\alpha_0) \ F \tag{16.25}$$

and notice that by multiplying both sides of (16.25) by $(k - f_0)^t$ and using (16.18) one obtains

$$\tau^2 \overline{\alpha_0} = (\chi/\alpha_0) \ (k - f_0)^t F \tag{16.26}$$

Now we reach a crossroads. The strategy behind our approach in Section (16.4) was to project (16.25) onto $H^{2\perp}_{N,0}$. In this case it produces a Toeplitz operator type formula, $P_{H^{2\perp}_{N,0}}[\ \overline{\alpha_0}\,\overline{k} - \overline{\alpha_0}\,\overline{f}\ ] = 0$, i.e., $\mathcal{T}_{k-f}[\ \alpha\ ] = 0$.

The strategy in Nehari and Commutant Lifting approaches is to project both terms in (16.25) onto $H^2_{N,0}$. This yields

$$P_{H^2_{N,0}}[\ \overline{\alpha_0}\,\overline{k}\ ] = (\chi/\alpha_0) \ F \tag{16.27}$$

Multiplying both sides of (16.27) by $(k - f)^t$, using (16.26), and taking conjugates of both sides we have

$$(\overline{k - f})^t P_{H_N^{2\perp}}[\, \alpha_0 k \,] = \tau^2 \alpha_0 \tag{16.28}$$

Taking projection of both sides of (16.28) onto H_N^2 gives

$$P_{H_N^2}[\, \overline{k}^t P_{H_N^{2\perp}}[\, \alpha_0 k \,]\,] = \tau^2 \alpha_0 \tag{16.29}$$

Now equation (16.29) is precisely the Nehari - Commutant Lifting formula (16.6).

We see that the Nehari- Commutant Lifting formula is remarkable in that it puts together, in a single equality, both relations (16.18) and (16.19), while at the same time eliminates f and F (our approach in Section 3 eliminates τ and F). It is also notable that all this is accomplished with a formula that is *linear* in α. It is unfortunate that these properties do not hold for functions Γ that are more general than (16.3).

16.6 References

[A] Ando

[AAK1] V. M. Adamjam, D. Z. Arov and M. G. Krein. Infinite Hankel matrices and generalized problems of caratheodory-Fejer and F. Riesz, Functional Anal. Appl. 2 (1968), pp. 1-18.

[AAK2] V. M. Adamjam, D. Z. Arov and M. G. Krein. Analytic properties of Schmidt paris for Hankel operator and the generalized Scur-Takagi problem. Math. USSR-Sb. 15 (1972), pp. 15-78.

[AAK3] V. M. Adamjam, D. Z. Arov and M. G. Krein. Infinite block Hankel matrices and related extension problems, Amer. Math. Soc. Translations 111 (1978), 133-156.

[BHM] J. Bence, J. W. Helton and D. E. Marshall. “Optimization over H-infinity,” Proc. Conference on Decision and Control, Athens, Greece, December 1986.

[FFHKP] C. Foias, B. Francis, J. W. Helton, H. Kwakernaak, J.B.Pearson. H^∞ -control Theory. Lecture Notes in Mathematics No. 1496. E. Mosca and L. Pandolfi, editors. Springer Verlag, Berlin. 1991.

[HMer1] A Novel Approach to Accelerating Newton's Method for sup-norm optimization arising in H^∞-control. J. Optim. Theory Apps. 1993

[HMer2] J. W. Helton and O. Merino. Conditions for Optimality over $\mathcal{H}^\infty$. SIAM J. Cont. Opt., Vol. 31, No. 6, November 1993.

[HMW] J. W. Helton, O. Merino and T. E. Walker. Algorithms for Optimizing over Analytic Functions. Indiana U. Math. J., Vol 42, No.3 (1993) .

[NF] Sz.-Nagy, B. and Foias, C. Harmonic Analysis of Operators on Hilbert Space. North Holland, 1970.

[S] Sarason, D. Generalized interpolation in H^∞. Trans. Amer. Math. Soc. vol 127, pp. 179-203. 1967.

[SKL] S??, T. Kailath and L??

Chapter 17

A sufficient condition for optimality

A sufficient condition for optimality in the problem OPT is presented in this chapter. The main result is given in Section 17.1, but the proof is only given at the end of the chapter in Section 17.4. An example is given in Section 17.2, and some qualitative results on OPT from the literature are given in Section 17.3.

17.1 Complete characterization of solutions

We will use the following notation for the second order partial derivatives:

$$\frac{\partial^2\Gamma}{\partial\overline{z}\partial z}(e^{j\theta},z)\ \left(\text{resp. } \frac{\partial^2\Gamma}{\partial z\partial z}(e^{j\theta},z)\ \right) \tag{17.1}$$

to represent the $N \times N$ matrix-valued function with (ℓ,k) entries given by

$$\frac{\partial^2\Gamma}{\partial\overline{z_\ell}\partial z_k}(e^{j\theta},z)\ \left(\text{resp. } \frac{\partial^2\Gamma}{\partial z_\ell\partial z_k}(e^{j\theta},z)\ \right)$$

The next result states a sufficient condition for a function to be a solution to OPT.

Theorem 22 *Let Γ be of class C^3 and $f^* = (f_1,\ldots,f_N)$ in A_N be such that*

$$a(e^{j\theta}) := \left(\frac{\partial\Gamma}{\partial z_1}(e^{j\theta},f^*(e^{j\theta})),\ldots,\frac{\partial\Gamma}{\partial z_N}(e^{j\theta},f^*(e^{j\theta}))\right) \neq (0,\ldots,0), \qquad \forall\theta$$

Set $A(e^{j\theta}) = \frac{\partial^2\Gamma}{\partial z\partial\bar{z}}(e^{j\theta}, f^*(e^{j\theta}))$, *and* $B(e^{j\theta}) = \frac{\partial^2\Gamma}{\partial^2 z}(e^{j\theta}, f^*(e^{j\theta}))$.

If the function f^* *is a directional solution to* OPT *then, in addition to I and II of Theorem 20, the following condition holds:*

III *For every nonzero* $h \in \mathcal{N} = \{h \in A_N : a(e^{j\theta})^t h(e^{j\theta}) = 0, \forall e^{j\theta}\}$,

$$\sup_\theta \overline{h(e^{j\theta})}^t \mathcal{A}(e^{j\theta})h(e^{j\theta}) + Re[h(e^{j\theta})^t \mathcal{B}(e^{j\theta})h(e^{j\theta})] \geq 0$$

Conversely, if conditions I, II, and III with strict inequality hold, then f^* *is a directional minimizer.*

The interested reader may check the reference **[HMer2]** for details.

The following theorem replaces III with a (weaker) condition that is easier to check than III. When taken together with I and II, they are sufficient for optimality.

Theorem 23 *Let* Γ, f^*, a, $\mathcal{A}$ *and* $\mathcal{B}$ *be as in Theorem 20. If, in addition to I and II of Theorem 20, there exist* $\delta > 0$ *and* $e^{j\theta}$ *such that for every* $z \in C^N$,

$$a(e^{j\theta})^t z \;=\; 0 \;\Rightarrow\; \bar{z}^t\mathcal{A}(e^{j\theta})z + Re[z^t\mathcal{B}(e^{j\theta})z] \;\geq\; \delta \mid\mid z \mid\mid^2_{C^N} \qquad (17.2)$$

Then f^* *is a directional solution of OPT. Moreover, if the statement (17.2) holds for some* $\delta > 0$ *and all* $e^{j\theta} \in \partial D$ *and* $z \in C^N$, *then* f^* *is a strict local optimizer.*

We present in the Section 17.4 the proof of Theorem 23 since it has not been published before. The reader may skip it without harm. We have the following corollaries of Theorem 23:

Corollary 1 *If* f^* *satisfies I and II, and if the sublevel sets of* Γ *are strictly pseudo-convex, then* f^* *is a strict local optimizer.*

Corollary 2 *If* f^* *satisfies I and II, and if* $\Gamma(e^{j\theta}, z)$ *is strictly convex in* z, *then* f^* *is a strict local optimizer.*

17.2 An Example, continued

We continue our study of OPT for the performance function in (15.3). We will show that f_1 and f_2 are indeed local solutions, and that f_0 is not.

We begin by verifying that f_1 and f_2 are local solutions. To use Theorem 23 we need the second order partial derivatives:

$$\frac{\partial^2\Gamma}{\partial z \partial \bar{z}} = \begin{pmatrix} \begin{array}{c} 1 + |e^{i\theta} + 0.1 + 0.1z_2|^2 \\ +|0.1z_2 + 0.1|^2 \end{array} & \begin{array}{c} (0.1\bar{z}_1 + 0.1)(e^{i\theta} + 0.1 + 0.1z_2) \\ +(e^{-i\theta} + 0.1\bar{z}_1 + 0.1)(0.1 + 0.1z_2) \end{array} \\ \begin{array}{c} (0.1\bar{z}_2 + 0.1)(e^{i\theta} + 0.1 + 0.1z_1) \\ +(e^{-i\theta} + 0.1\bar{z}_2 + 0.1)(0.1 + 0.1z_1) \end{array} & \begin{array}{c} 1 + |e^{i\theta} + 0.1 + 0.1z_1|^2 \\ +|0.1z_1 + 0.1|^2 \end{array} \end{pmatrix}$$

and

$$\frac{\partial^2\Gamma}{\partial^2 z} = \begin{pmatrix} 0 & \begin{array}{c} 0.1(200 + e^{-j\theta}\bar{z}_1 + e^{-j\theta}\bar{z}_2 \\ +0.2(\bar{z}_1 + \bar{z}_2 + \bar{z}_1\bar{z}_2)) \end{array} \\ \begin{array}{c} 0.1(200 + e^{-j\theta}\bar{z}_1 + e^{-j\theta}\bar{z}_2 \\ +0.2(\bar{z}_1 + \bar{z}_2 + \bar{z}_1\bar{z}_2)) \end{array} & 0 \end{pmatrix}$$

With the functions $\mathcal{A}$ and $\mathcal{B}$ defined as in Theorem 22, we have that

$$\mathcal{A}(e^{j\theta}) = \begin{pmatrix} 18.82 - 5.8cos(\theta) & 3.1e^{j\theta} - 2.9\overline{e^{j\theta}} - 17.98 \\ 3.1\overline{e^{j\theta}} - 2.9e^{j\theta} - 17.98 & 21.22 + 6.2cos(\theta) \end{pmatrix}$$

$$\mathcal{B}(e^{j\theta}) = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$$

If θ is fixed, then clearly $a(e^{j\theta})^t z = 0$ when z has the form $z^t = (z_1, -z_1) = z_1(1, -1)$ for some $z_1 \in C$. For such z we have that

$$\begin{aligned} \overline{z}^t\mathcal{A}(e^{j\theta_0})z + & \quad Re[z^t\mathcal{B}(e^{j\theta_0})z] \\ = & \quad (1,-1)A(e^{j\theta})\begin{pmatrix} 1 \\ -1 \end{pmatrix}|z_1|^2 + Re[(1,-1)B(e^{j\theta})\begin{pmatrix} 1 \\ -1 \end{pmatrix}z_1^2] \\ = & \quad 72|z_1|^2 + Re[-4z_1^2] \\ \geq & \quad 68|z_1|^2 \\ = & \quad 34|z|^2 \end{aligned} \tag{17.3}$$

Therefore condition (17.2) holds for all θ and all z, and this implies that f_1 is a strict local minimizer.

Now we show that $f_0(e^{j\theta}) = (0,0)$ is not a directional minimizer (in particular, it is not a local minimizer). We claim that the function $h(e^{j\theta}) = (1,-1)^t$ is a descent direction relative to f_0, thus implying that f_0 is not a directional minimizer. Indeed, from

$$\begin{aligned}\Gamma(e^{j\theta}, f_0(e^{j\theta}) & +th(e^{j\theta})) \\ & = \Gamma(e^{j\theta}, t, -t) \\ & = 2t^2 + |100 + te^{j\theta} - 0.1t^2|^2 + |100 - te^{j\theta} - 0.1t^2|^2 \\ & = 20000 - 36t + 0.02t^4\end{aligned}$$

we see that for all small $t > 0$,

$$\sup_\theta \Gamma(e^{j\theta}, f_0(e^{j\theta}) + th(e^{j\theta})) < 20000$$

17.3 Existence, smoothness, and uniqueness of solutions

In this section we return to basic qualitative questions which were addressed for $N = 1$ in Section 10.4. Do solutions in H_N^∞ exist? Are solutions unique? Are solutions smooth?

Again, $S_\theta(c)$ refers to the sublevel sets

$$S_\theta(c) = \{z : \Gamma(e^{j\theta}, z) \leq c\}$$

The main difference between the $N = 1$ and $N > 1$ case is that in higher dimensions solutions may not be unique. We saw an example of non uniqueness earlier in Section 15.4. Those steeped in the lore of Several Complex Variables will appreciate the fact that Γ in this example is strictly plurisubharmonic in z.

Our standard assumption on the (OPT) problem is:

SA Γ depends smoothly on θ, is real analytic in z (and in $\bar{z}$) and has gradient $\overline{\frac{\partial \Gamma}{\partial z}(e^{i\theta}, z)}$ which never vanishes when $\Gamma(e^{i\theta}, z) = \gamma^*$. The sets $\mathcal{S}_\theta(\gamma^*)$ are connected, simply connected, have nonempty interior, and are uniformly bounded in θ.

While γ^* may not be known in advance in a particular situation, one might verify that all $\mathcal{S}_\theta(\gamma)$ for a wide range of γ satisfy these conditions; this is because the conditions are not very restrictive.

We now give a list of results. Many of these can be found in **[HH]**, **[HMar]**, **[HMer2]**. Definitions of the more specialized terms in the theorems are given below.

When $N \geq 1$ we have

Theorem 24 *If* $\mathcal{S}_\theta(\gamma^*)$ *is strictly convex (uniformly in* θ*), then an* H^∞_N *solution* f^* *to* OPT *exists and it is unique. Also* $f^*(e^{i\theta}) \in \partial\mathcal{S}_\theta$ *for a.e.* θ*.*

Theorem 25 *Suppose* SA *holds and that each* $\mathcal{S}_\theta$ *is polynomially convex. Then an* H^∞_N *solution* f^* *to* OPT *exists. Moreover, if a sequence* $f_k \in H^\infty_N$ *approximately solves* OPT *(in the sense* $\sup_\theta \Gamma(e^{i\theta}, f_k(e^{i\theta})) = \gamma^k$ *with* $\gamma^k \searrow \gamma^*$*), then a subsequence which converges in normal family sense has as its limit a function* f_∞ *in* H^∞_N *which satisfies* $\Gamma(e^{i\theta}, f_\infty(e^{i\theta})) \le \gamma^*$ *almost everywhere.*

A strictly convex set is a convex set with no line segment contained in its boundary. The unit ball in C^N is strictly convex, while the unit ball in the space M_{mn} of $m \times n$ matrices is not unless $m = 1$ or $n = 1$. Polynomially convex sets are a much broader class of sets than convex sets. While convex sets are intersections of half spaces $\{w : Re\ell(w) \le 0\}$ where ℓ is a linear function into the complex plane. While *polynomially convex sets* are intersections of sets of the form $\{w : Rep(w) \le 0\}$ where p is a polynomial.

17.4 Proof of Theorem 23

The first part of Theorem 23 is a consequence of the follwing lemma:

Lemma 3 *The hypothesis of Theorem 23 implies III with strict inequality.*

Note that since the mapping given by

$$H(e^{j\theta}, z) = \overline{z}^t \mathcal{A}(e^{j\theta})z + Re[z^t \mathcal{B}(e^{j\theta})z] \tag{17.4}$$

is continuous in z and $e^{j\theta}$jointly, there exists an arc $J \subset \partial D$ centered at $e^{j\theta}$ such that relation (17.2) holds for all $e^{j\theta} \in J$. If now III does not hold for some nonzero $h \in \mathcal{N}(a)$, then (17.2) implies that this function satisfies $h(e^{j\theta}) = 0$ for all $e^{j\theta} \in J$. Since h is analytic, h is the function 0, contradicting the assumption on the choice of h. ■

To prove the second assertion in Theorem 23, the following two lemmas are needed.

Lemma 4 *Assume the notation and hypotheses of Proposition 23. If* $h \in A_n$ *is such that* $a(e^{j\theta})^t h(e^{j\theta})$ *is not the function* 0 *then*

$$\sup_\theta Re[a(e^{j\theta})^t h(e^{j\theta})] > 0$$

Proof of Lemma 4. Let $F = (F_1, \ldots, F_N)^t$ and $\lambda(e^{j\theta})$ be the functions that appear in statement II. The scalar function $g(\zeta) = \zeta F(\zeta)^t h(\zeta)$ is analytic and also belongs to the Hardy space H^1. Since $g(0) = 0$ and g is an open mapping (g is nonconstant), it maps a neighborhood of 0 onto a neighborhood of 0. Hence for some $\zeta \in D$, we have that $g(\zeta) > 0$. The Poisson representation of g implies that $Re[g(e^{j\theta})] > 0$ on a set $J \subset \partial D$ with positive measure. Then

$$\sup_\theta Re[a(e^{j\theta})h(e^{j\theta})] = \sup_\theta \lambda(e^{j\theta}) Re[g(e^{j\theta})] > 0 \qquad \blacksquare$$

Lemma 5 *Assume the notation of Proposition 23. There exists* η_0 *such that if* $|Re[a(e^{j\theta})^t z| < \eta_0 ||z||^2_{C^N}$*, then* $H(e^{j\theta}, z) \geq \delta/2$.

Proof of Lemma 5. This is a consequence of the continuity of H. $\blacksquare$

Assume for the sake of contradiction that $\{h_n\}$ is a sequence in A_N such that $||h_n||_\infty \to 0$ as $n \to \infty$ and

$$\sup_\theta \Gamma(e^{j\theta}, f^*(e^{j\theta}) + h_n(e^{j\theta})) \leq \sup_\theta \Gamma(e^{j\theta}, f^*(e^{j\theta})) \quad \forall n \tag{17.5}$$

For each n choose θ_n to be a point at which $\sup_\theta Re[a(e^{j\theta})^t h_n(e^{j\theta})]$ is attained, that is,

$$Re[a(e^{j\theta_n})h_n(e^{j\theta_n})] = \sup_\theta Re[a(e^{j\theta})^t h_n(e^{j\theta})] \tag{17.6}$$

We now consider three cases.

Case 1. *The following inequality holds for an infinite number of n's:*

$$0 < Re[a(e^{j\theta_n})^t h_n(e^{j\theta_n})] < \eta_0 ||h_n(e^{j\theta_n})||_{C^N} \tag{17.7}$$

Drop terms if necessary to assume that (17.7) holds for all n. Note that $h_n(e^{j\theta_n}) \neq 0$. We now use the Taylor expansion of Γ in the variable z to define

$$\begin{aligned} x_n := \; & \Gamma(e^{j\theta_n}, f^*(e^{j\theta_n}) + h_n(e^{j\theta_n})) - \Gamma(e^{j\theta_n}, f^*(e^{j\theta_n})) \\ & -2Re[a(e^{j\theta_n})^t h_n(e^{j\theta_n})] - H(e^{j\theta_n}, z)[h_n(e^{j\theta_n}), h_n(e^{j\theta_n})] \end{aligned}$$

By (17.6) and (17.5) we have that

$$x_n \leq -H(e^{j\theta_n}, h_n(e^{j\theta_n})) \tag{17.8}$$

Note that (17.7) allows us to combine (17.8) with Lemma 5 to conclude that for all n,

$$\frac{x_n}{||h_n(e^{j\theta_n})||_{C^N}^2} \leq -\frac{\delta_0}{2} \tag{17.9}$$

On the other hand, since $h_n(e^{j\theta_n}) \longrightarrow 0$ as $n \longrightarrow \infty$, and by the smoothness of Γ, we have that

$$\frac{x_n}{||h_n(e^{j\theta_n})||_{C^N}^2} \longrightarrow 0 \text{ as } \ n \longrightarrow \infty \tag{17.10}$$

which contradicts (17.9).
Case 2. *For all n the following relation holds:*

$$Re[a(e^{j\theta_n})^t h_n(e^{j\theta_n})] \geq \eta_0 ||h_n(e^{j\theta_n})||_{C^N} \tag{17.11}$$

Set

$$y_n := \Gamma(e^{j\theta_n}, f^*(e^{j\theta_n}) + h_n(e^{j\theta_n})) - \Gamma(e^{j\theta_n}, f^*(e^{j\theta_n})) - 2Re[a(e^{j\theta_n})^t h_n(e^{j\theta_n})]$$

Since $h_n(e^{j\theta_n}) \longrightarrow 0$ as $n \longrightarrow \infty$, and by the smoothness of Γ we have,

$$\frac{y_n}{||h_n(e^{j\theta_n})||_{C^N}} \longrightarrow 0 \text{ as } \ n \longrightarrow \infty \tag{17.12}$$

But by relation (17.11) and by the hypothesis (17.5), we have that

$$\frac{y_n}{||h_n(e^{j\theta_n})||_{C^N}} \leq -\frac{\delta_0}{2} \tag{17.13}$$

which contradicts (17.12).
Case 3. *The following equality holds for an infinite number of n's:*

$$Re[a(e^{j\theta_n})^t h_n(e^{j\theta_n})] = 0 \tag{17.14}$$

Drop terms if necessary to assume that this relation holds for all n. Lemma 4 implies that $Re[a(e^{j\theta})^t h_n(e^{j\theta})] = 0$ for all $e^{j\theta}$ and all n. Define x_n as in Case 1 but now θ_n is chosen so that $||h(e^{j\theta})||_{C^N}$ attains its maximum at $\theta = \theta_n$. Just as before, the smoothness of Γ and the fact that $h_n(e^{j\theta_n}) \longrightarrow 0$ as $n \longrightarrow \infty$ imply a relation similar to (17.10). On the other hand, the hypotheses (17.14) and (17.2) imply a statement similar to (17.9) but with δ instead of $\delta/2$, thus yielding a contradiction. ∎

17.5 References

The main theorem in this section is from **[HMer2]**. More results can be found in **[H3]** and in **[HMar]**. See **[H1]** for connections with other areas of Mathematics. Several Complex Variables connections are discussed in **[HMer1]**, see also **[HH]**, **[HMar]** and **[HV]**.

[H1] J. W. Helton. Operator Theory, Matrices and Electrical Engineering. NSF. 1987

[H2] J. W. Helton. "Optimization over spaces of analytic functions and the Corona problem," J. Operator Theory, 15 (1986), No. 2, 359–375.

[H3] J. W. Helton: Optimal frequency domain design vs. an area of several complex variables, plenary address. MTNS, 1989.

[HH] J. W. Helton and R. Howe. "A bang-bang theorem for optimization over spaces of analytic functions," J. Approx. Theory, 47 (1986), No. 2, 101–121.

[HMar] J. W. Helton and D. Marshall: "Frequency domain design and analytic selections," Indiana Univ. Math. Journal, 39 (Spring 1990), No. 1, 157–184.

[HMer1] J. W. Helton and O. Merino: "Optimal Analytic Disks," Collection: Several complex variables and complex geometry, Part 2, (Santa Cruz, CA, 1991) Proc. Sympos. Pure Math., 5 251–262.

[HMer2] J. W. Helton and O. Merino. Conditions for Optimality over $\mathcal{H}^\infty$. SIAM J. Cont. Opt., Vol. 31, No. 6, November 1993.

[HV] J. W. Helton, and A. Vityaev. Preprint

Part V

Appendices

Appendix A

History and Perspective

While historians trace control to Archimedes time and the theory of control to a paper of James Clerk Maxwell on governors, the subject called classical control began in the second world war in labs in England and the U.S. which designed radar-driven antiaircraft guns. The beginnings of classical control were summarized in a book **[JNP]** by James, Nichols, and Phillips. According to Ralph Phillips, a book which was very influential in their lab was Bode's famous book on amplifiers. However, this book was not referenced in **[JNP]**. The primary technique which emerged was adjusting parameters in low order (e.g., degree two) rational functions and checking that the graphs lie in certain regions. Classical control dominated industrial practice for many years, even though it could only be taught by example and its main technique was trial and error.

Much of the theory of control in the 1960's and 70's focused on achieving desired frequency domain performance as closely as possible in a mean-square-error sense. We have nothing to add to the literature here so do not give a historical treatment.

The subject of optimizing worst case error in the frequency domain along its present lines started not with control but with circuits. One issue was to design amplifiers with maximum gain over a given frequency band. Another was the design of circuits with minimum broadband power loss. Indeed H^∞ control is a subset of a broader subject, worst case design in the frequency domain. In paradigm engineering problems this produces what a mathematician calls an 'interpolation problem' for analytic functions. The techniques of Nevanlinna-Pick interpolation had their first serious introduction into engineering in a SISO circuits paper of Youla and Saito **[YS]** in the middle 1960's. Further development waited until the mid-seventies, when Helton **[H1–H3]** applied them and more general techniques from operator theory to amplifier problems. The methods of commutant lifting **[A, NF, S]** and of Admajan-Arov-Krein **[AAK1–AAK3]** were used to solve MIMO optimization problems. The disk method used here was first described in an engineering article on gain equalization **[H3]**, Tnd followed the foundational theory which was published in the mathematics article **[H2]** two years earlier.

In the late 1970s G. Zames **[Z]** began to marshal arguments indicating that H^∞ rather than H^2 was the physically proper setting for control. Zames suggested on several occasions that these methods were the appropriate ones for codifying classical control. These efforts yielded a mathematical problem which Helton identified as an interpolation problem solvable by existing means (see acknowledgement in **[ZF1]**). Zames and Francis **[ZF]** used this to solve the resulting problem. Soon thereafter Chang-Pearson **[CP]** and Francis-Helton-Zames solved it for many-input, many-output MIMO systems **[FHZ]**.

The pioneering work of Zames and Francis treated only sensitivity optimization. In 1983 three independent efforts emphasized bandwidth constraints, formulated the problem as a precise mathematics problem and indicated effective numerical methods for their solution: Doyle **[D1]**, Helton **[H4]**, and Kwakernaak **[K1]**. Helton treated the problem graphically, just as it is done in Part I of this book, by approximating multiple disk constraints with a single moving disk. Kwakernaak gives an equivalent method, but it is expressed algebraically in terms of two weight functions W_1 and W_2 on the sensitivity function $I - T$ and on T. Also all of these papers described quantitative methods which were soon implemented on computers. It was these papers actually laid out precisely the tradeoff in control between performance at low frequency and roll off at higher frequency and how one solves the resulting mathematics problem. This is in perfect analogy with amplifier design where one wants large gain over as wide a band as possible, producing the famous gain bandwidth trade off. Rather remarkable in textbook treatments of control through the 1970's was the absence (or cursory treatment) of bandwidth constraints. On the other hand, many control practitioners of that period would say that bandwidth constraints were important, but when they went to formulate them as a mathematical optimization problem would provide no high frequency constraint. Thus they made the fundamental mistake of H^∞ control (see Section 11.4), and the concomitant confusion cost the methods much credibility. As of the 1990's the control community's views have become more realistic. This conceptual shift is one of the main accomplishments of H^∞ control. The moral of the story is that nothing stabilizes a nebulous philosophical discussion like the the ability to solve the math problems which arise exactly. Mush dries up quickly under this torch.

All of the traditional methods in H^∞ solved only optimization problems where the specification sets $\mathcal{S}_\theta$ are disks. There is a beautiful and extensive theory of this. Much of the mathematical theory came directly from existing work in the study of operators on Hilbert Space. However, an important accomplishment of theoretical engineers was to formulate the problem in a new set of coordinates (called state space coordinates) which are well suited to many problems. This is an extremely elegant theory (cf [**BGR**], [**F**], [**GL**], [**GM**]).

As basic as it is, the work on internal stability treated in Part I was done only recently. In the famous article of Youla, Jabr, Bongiorno [**YJB**], the concept of internal stability was introduced and reduced to a mathematical problem in which all H^∞ functions meet a given interpolation constraints. The authors then solved the H^2 control problem, (rather than optimizing over H^∞ as is done in this book).

Independent of the rise of H^∞ control was the Horowitz approach [**Ho**]. The relationship between his approach and ours is described in [**BHMer**]. Another independent development was Tannenbaum's [**T**] very clever use of Nevanlinna-Pick interpolation in a control problem in 1980. Also early on the H^∞ stage was Kwakernaak's polynomial theory cf.[**K2**]. Another major development that dovetailed closely with the invention of H^∞ control was a tractable theory of plant uncertainty. A good historical treatment appears in [**DFT**]. Another application of these techniques is to robust stabilization of systems ([**Kim**])

The mathematical theory in this book (Part IV) is quite different and more general than the classical mathematical theory (Nevanlinna-Pick, Nehari) which applies only to disk problems. In Chapter 16 we showed how the basic result of Nehari's for disks follows from our approach. Surprisingly, it is not much more difficult to derive by our approach than it is classically. The reader interested in connections of the theory developed here with other areas may want to see [**Dy**], [**FF**] and [**H8**]. Work on the general theory for solving the optimization problem OPT described in Part III and Part IV started in 1981 by Helton and Howe, who worked on convex problems [**HH**]. Since then theoretical work has been done by J. W. Helton, S.Hui, D.Marshall, O. Merino, Z.Slodkowski, A. Vityaev K. Lenz and E. Wegert in papers listed in the bibliography. Theoretical work pertaining to numerical issues has been continued by Helton, Merino, and T. Walker, [**HMer2**], [**HMer3**], [**HMW**], and numerical testing and more theoretical work has been done by J. Bence, J. W. Helton, O. Merino, J. Myers, D. Schwartz, T. Walker. Interestingly, the theory developed to solve the engineering problem OPT has strong ties with ongoing work in the study of functions of several complex variables. For an account of this see [**H5**], [**HMer1**]. The main result used here (Theorem [**OPT** $N > 1$])

is taken from **[HMer2]**. The paper **[HMW]** introduced the Newton computer algorithm of Chapter 7. The reason Newton's method had not been succesfull before on such an old problem is probably because sup norm is not a smooth performance function. Thus one needs a special approach to discover it.

The SISO winding number results PART IV are from **[H6]**.

References

[AAK1] V. M. Adamjam, D. Z. Arov and M. G. Krein. Infinite Hankel matrices and generalized problems of caratheodory-Fejer and F. Riesz, Functional Anal. Appl. 2 (1968), pp. 1-18.

[AAK2] V. M. Adamjam, D. Z. Arov and M. G. Krein. Analytic properties of Schmidt pairs for Hankel operator and the generalized Schur-Takagi problem. Math. USSR-Sb. 15 (1972), pp. 15-78.

[AAK3] V. M. Adamjam, D. Z. Arov and M. G. Krein. Infinite block Hankel matrices and related extension problems, Amer. Math. Soc. Translations 111 (1978), 133-156.

[A] T. Ando. On a pair of commuting contractions. Acta Sci. Math. (Szeged) 24, 1963, 88-90.

[BGR] J. A. Ball, I. Gohberg and L. Rodman. Interpolation of rational matrix functions. Birkhauser-Verlag, Basel, 1990.

[BHM] J. Bence, J. W. Helton and D. E. Marshall. "Optimization over H-infinity," Proc. Conference on Decision and Control, Athens, Greece, December 1986.

[BHMer] F. N. Bailey, J. W. Helton and O. Merino: "Alternative approaches in frequency domain design of single loop feedback systems with plant uncertainty," to appear (1994).

[CP] B. C. Chang and B. Pearson Jr., Optimal disturbance reduction in linear multivariable systems. IEEE Trans. Automat. Contr., vol AC-29, pp. 880-887, 1984.

[D1] J. C. Doyle. Synthesis of robust controllers and filters. Proceedings of 22nd. IEEE Conference on Decision and Control, San Antonio, Texas 1983.

[D2] J. C. Doyle. Robust stability with structured perturbations. Proc. Sym. on Mathematical Theory of Networks and Systems. Stockholm, 1985.

[DFT] Doyle, J. , Francis, B. A., and Tannenbaum, A. Feedback Control Theory. Macmillan Pub. Co., New York, 1992.

[DGKF] J. C. Doyle, K Glover, P. P). Khargonekar, B.A. Francis. State-space solutions to standard H^2 and H^∞ control problems. IEEE Trans. on Automat. Contr. vol 34, pp. 831-847, 1989.

[DS] Doyle, J. and G. Stein, Multivariable feedback design: Concepts for a classical/modern synthesis. IEEE Trans. Automat. Contr., vol. AC-26, pp. 4-16, 1981.

[Dy] H. Dym. J contractive matrix functions, reproducing kernel Hilbert spaces and interpolation. CBMS Reg. Conf. Ser. in Math., vol 71, Amer. Math. Soc., Providence, RI, 1989

[F] B. Francis. First course in H^∞-control. Springer Verlag. 1987

[FHZ] Francis, B. A., Helton, J. W. and Zames, G. H^∞-optimal feedback controller for linear multivariable systems. IEEE Trans. Auto Comnt. , vol AC-29, pp. 888-900.

[FFHKP] C. Foias, B. Francis, J. W. Helton, H. Kwakernaak, J.B.Pearson. H^∞ -control Theory. Lecture Notes in Mathematics No. 1496. E. Mosca and L. Pandolfi, editors. Springer Verlag, Berlin. 1991.

[FF] C. Foias and Frazho. The commutant lifting approach to interpolation problems. Birkhauser-Verlag, Basel, 1991.

[GL] K. Glover and D. C. McFarlane. Robust Controller Design Using Normalized Coprime Factor Pland Descriptions. Vol. 138 in Lecture Notes in Control and Information Sciences. Springer Verlag, New York. 1990.

[GM] M. Green and D. J. N. Limebeer. Book

[H1] J. W. Helton. "Operator theory and broadband matching," Proceedings of Eleventh Annual Allerton Conference on Circuits and Systems Theory, 1976.

[H2] J. W. Helton. "A mathematical view of broadband matching," IEEE International Conference on Circuits and Systems Theory, New York, 1978.

[H3] J. W. Helton. "Broadbanding gain equalization directly from data," IEEE Trans. Circuits and Systems, CAS-28 (1981), No. 12, 1125–1137.

[H4] J. W. Helton. "An H-infinity approach to control," IEEE Conference on Decision Control, San Antonio, Texas, December 1983

[H5] J. W. Helton. "Worst case analysis in the frequency domain: The H-infinity approach to control," IEEE Trans. Auto. Cont., AC-30 (December 1985), No. 12, 1154–1170.

[H6] J. W. Helton. "Optimization over spaces of analytic functions and the Corona problem," J. Operator Theory, 15 (1986), No. 2, 359–375.

[H7] J. W. Helton: Optimal frequency domain design vs. an area of several complex variables, plenary address. MTNS, 1989.

[H8] J. W. Helton. Operator Theory, Analytic Functions, Matrices and Electrical Engineering. CBMS regional conference series in Mathematics. No. 68, Amer. Math. Soc., Providence, RI. 1987.

[HH] J. W. Helton and R. Howe. "A bang-bang theorem for optimization over spaces of analytic functions," J. Approx. Theory, 47 (1986), No. 2, 101–121.

[HMar] J. W. Helton and D. Marshall: "Frequency domain design and analytic selections," Indiana Univ. Math. Journal, 39 (Spring 1990), No. 1, 157–184.

[HMer1] J. W. Helton and O. Merino: "Optimal Analytic Disks," Collection: Several complex variables and complex geometry, Part 2, (Santa Cruz, CA, 1991) Proc. Sympos. Pure Math., 5 251–262.

[HMer2] J. W. Helton and O. Merino. Conditions for Optimality over $\mathcal{H}^\infty$. SIAM J. Cont. Opt., Vol. 31, No. 6, November 1993.

[HMer3] A Novel Approach to Accelerating Newton's Method for supnorm optimization arising in H^∞-control. J. Optim. Theory Apps. 1993

[HMW] J. W. Helton, O. Merino and T. E. Walker. Algorithms for Optimizing over Analytic Functions. Indiana U. Math. J., Vol 42, No.3 (1993) .

[HV] J. W. Helton, and A. Vityaev. Preprint

[Ho] Horowitz, I. M. Synthesis of Feedback Systems. Academic Press, NY. 1963.

[Hui] S. Hui. Qualitative Properties of Solutions to H^∞-Optimization. Problems. J. of Functional Analysis **75** (1987) 323,348.

[JNP] James, H. M., Nichols N. B. and Phillips, R. P. Theory of Servomechanisms, Radiation Lab. Series, Vol 25, McGraw-Hill, New York, 1947.

[Kim] H. Kimura Robust stabilizability for a class of transfer functions, IEEE Trans Auto. Control AC-29 1984 788-793

[K1] H. Kwakernaak Robustness Optimization of Linear Feeedback Systems. IEEE Conf. Decision Control, San Antonio, Texas, December 1983

[K2] H. Kwakernaak. EEC 1993

[L] K. Lenz. Properties of certain optimal weighted mixed sensitivity designs. To appear.

[NF] Sz.-Nagy, B. and Foias, C. Harmonic Analysis of Operators on Hilbert Space. North Holland, 1970.

[S] Sarason, D. Generalized interpolation in H^∞. Trans. Amer. Math. Soc. vol 127, pp. 179-203. 1967.

[T] A. Tannenbaum. Feedback stabilization of plants with uncertainty in the gain factor. Int. J. of Control, vol. 32, pp1-16, 1980.

[W] E. Wegert Nonlinear Boundary value Problems for Holomorphic Functions and Singular Integral Equations, Academie Verlag, Berlin 1992

[YJB1] D.C. Youla, H. A. Jabr, J.J. Bongiorno. Modern Wiener-Hopf design of optimal controllers -Part I: The single input - single output case. IEEE Trans. Automat. Contr., vol AC-21, pp. 319-338, 1976.

[YJB2] D.C. Youla, H. A. Jabr, J.J. Bongiorno. Modern Wiener-Hopf design of optimal controllers -Part II: The multivariable case. IEEE Trans. Automat. Contr., vol AC-21, pp. 3-13, 1976.

[YS] D.C. Youla and M. Saito. Interpolation with positive real functions. H. Franklin Inst., vol 284, pp. 77-108, 1967.

[Z] Optimal sensitivity and feedback: weighted seminorms, approximate inverses, and plant invariant schemes. Proc. Allerton Conf. 1979.

[ZF1] G. Zames and B. Francis. Feedback and minimax sensitivity. Advanced Group for Aerospace Research and Development, NATO Lecture Notes, No. 117, Multivariable Analysis and Design Techniques

[ZF2] G. Zames and B. Francis. Feedback, minimax sensitivity, and optimal robustness, IEEE Trans. Automat. Contr., vol AC-28, PP. 585-601, 1983

Appendix B

Other Software

Appendix C

Pure Mathematics and Problems

We have taken the viewpoint of engineering and computation in several parts of the book. Much of it could be looked at simply as mathematics. One implication of this work is that it suggests a new class of problems in functional analysis and operator theory or in several complex variables if that is your viewpoint.

First we emphasize that a pure mathematician might read Part III, then follow up with part IV. One who despises computation might read the how to run the program part quickly, but the parts which describe numerical algorithms actually could have been cast in much more mainstream mathematical terms. To describe this, first focus on a statement which formulates what we have done to the old paradigm Nehari problem.

Claim.

For the Nehari problem of approximating a given function $k \in L^\infty$ *by functions* $f \in H^\infty$,

$$\inf_{f \in H^\infty} ||k - f||_\infty,$$

there are two different sets of optimality conditions for solutions f *each having the property that their associated differential is invertible at almost all locations.*

1. *Equations (16.5) and (16.6) in Theorem 21 due to Nehari.*
2. *The key optimality condition* $T(f, \beta) = 0$ *determined by equations (13.16) and (13.18) in Chapter 13), or equations (16.8) and (16.9) in Chapter 16.*

We call this a claim rather than the theorem because the differentials are operators on spaces of functions and we did not say exactly what the spaces are.

Idea of Proof. Here we shall denote the Nehari-Commutant Lifting conditions (16.5) and (16.6) in Theorem 21 as a system of equations

$$\mathcal{N}(\alpha, \lambda) = 0$$

where $\lambda = \tau^2 > 0$ and $\alpha \in H^2_N$. The differential

$$\mathcal{N}'_{\alpha,\lambda} \begin{pmatrix} \varepsilon \\ \delta \end{pmatrix} := (R_k - \lambda)\varepsilon - \delta\alpha$$

where $R_k : H^2_N \longrightarrow H^2_N$ is defined by $R_k \varepsilon := P_{H^2_N} \left[\overline{k}^t P_{H^{2\perp}_N} k \varepsilon \right]$. for each $\varepsilon \in L^2_N$. Now R_k is a selfadjoint operator on L^2_N and λ is an eigenvalue of R_k which for almost all k has multiplicity one. Thus range of $R_k - \lambda$ has codimension one, so it is not invertible. Fortunately, the map $\delta \longrightarrow \delta\alpha$ has one dimensional range in L^2_N and since α is the null vector of $R_k - \lambda$ it is perpendicular to range $(R_k - \lambda)$. Therefore, range $\mathcal{N}^1_{\alpha,\lambda}$ equals L^2_N.

The optimality conditions for the OPT problem after projecting are denoted $T(f, \beta) = 0$ in chapters 13 and 16. The objective is to prove that $T'_{f,\beta}$ is invertible for almost all f, β which are solutions to OPT. The formula for $T'_{f,\beta}$ is given in Propostion 4 of Chapter 16. There are various spaces on which $T'_{f,\beta}$ could act. The easiest setting to analyze is when $T'_{f,\beta}$ acts from H^2_{N+1} to H^2_{N+1} (see **[HMW]**). Then it roughly has the form selfadjoint Toeplitz plus Hankel. The Hankel operator is compact and *the Toeplitz because of the special structure of* M_1 *can be shown to be invertible provided* wA *is positive definite and* $a(e^{j\theta}) = \Xi^n \lambda(e^{j\theta}) F(e^{j\theta})$ *where* F *is analytic inside the disk,* $\lambda \geq 0$ *and* $n \leq 1$. A is positive definite because Γ is strictly plurisubharmonic; indeed for the Nehari example $A = 1$. The optimality conditions guarantee that the condition on a holds and thus for "most" choices of symbol k we have invertibility. See Theorem 6 **[HMW]**.
■

Unfortunately $H^2_N + 1$ is the wrong space for analyzing Newton's method since a nonlinear map rarely maps $H^2_N + 1$ into itself. Thus we did the analysis of invertibility in the space of C^∞ functions with a topology imposed by a family of Sobolev norms. The proof and precise formulation can be found in **[HMW]**. ■

The technique applied here makes sense for many optimization problems in function theory. Find optimality conditions $\mathcal{C}$ for the problem. These are a set of equations involving primal and dual variables. In our problem and probably in many others. The differential of $\mathcal{C}$ is not invertible. Find some way to modify $\mathcal{C}$ and produce new optimality conditions $\widetilde{\mathcal{C}}$ with invertible [1]left invertible will do differential $\widetilde{\mathcal{C}}'$. Thus we have the pure mathematics problem: find optimality conditions with invertible differential. Possibly this is a natural type of mathematical result to seek for many function theory problems.

What we are saying is hardly shocking from the viewpoint of optimization theory. The surprising thing is that for the class of problems we study one does not immediately get the invertibility condition. Also it is a bit surprising that such basic principles of optimization theory were only recently applied to the very old Nevanlinna-Pick-Nehari problem.

For perspective and to make sure that the reader interprets this speculation widely (wildly) enough we mention some of the more interesting work on optimization in several complex variables. This concerns N-P-N type problems on the polydisk. There is striking work of Agler which derives beautiful matrix inequality conditions for checking if a particular N-P interpolation problem is solvable but for a norm stronger than the supnorm. Agler's view implemented successfully several problems is that instead of doing supnorm optimization one modifies the norm in some way so that the answer to the problem can be gotten from an eigenvalue problem or something similar. A very elegant result of this type, due to Cotlar and Sadosky, converts the Nehari problem in the polydisk for a BMO type norm to an eigenvalue problem.

The suggestion here is that we liberalize our notion of what it means to answer a question in function theory beyond insisting that an answer must have the form of an eigenvalue problem. It seems like a very reasonable pursuit for function theory problems from the pure mathematics point of view is to find optimality conditions whose differential is invertible. While this seems like a considerable loosening of the eigenvalue condition if we insist that our conditions use only explicit formulas, then it may not be that liberal a condition. Only time will tell if these ideas extend in many other directions.

Also we mention that there are extremely pretty connections between the theory of analytic disks developed in this book and ongoing work in Several Complex Variables. The article **[HMer1]** describes one line of connections while the paper **[HV]** quickly indicates some others. Readers in the area of Several Complex Variables might find these articles interesting.

---

1 1

Appendix D

Uncertainty

D.1 Introduction

So far we have considered the system $\mathcal{S}$ where the plant P is known. However in most physical systems typically there is a certain amount of "plant uncertainty", i.e., lack of knowledge about the plant.

To explain what uncertainty is, we use the example of an airplane. We can model the control of an airplane with a system of differential equations. These equations are the result of an idealization of the airplane that invariably does not include some elements, either because we are not aware that they affect the airplane, or because we choose not to include them for simplicity and are deemed to be relatively unimportant. The differential equations have a series of parameters that have physical significance and whose values we can tell only approximately.

We may further simplify the model by linearizing the system of differential equations. The model is surely not perfect, indeed at high frequencies it is highly inaccurate. Also there are factors affecting the system, among which we have changes in the properties of the components or even malfunction of some components.

A common approach to treating uncertainty when doing system design is to assume that there is available a known *reference* or *nominal* plant P_0, and that the true plant P of the system is, in the frequency domain description,

$$P(j\omega) = P_0(j\omega) + \Delta_0 \tag{D.1}$$

where Δ_0 is unknown. We may think of P as the result of perturbing P_0.

Uncertainty can have a devastating effect on a system. For example, a stable system may cease to be stable after a perturbation of the plant. This definitely is something that we do not want under any circumstances, so we shall require that reasonable amounts of uncertainty should never make a system that we have designed unstable. Thus we insist that all feasible plants make the system *internally stable.* This is called *robust stability.*

Uncertainty also affects the performance of the system. Ideally, we would like to design the system with the property that when the plant is perturbed, the resulting system satisfies the performance requirements originally set. This must hole for any feasible perturbation of the plant. Sometimes this is referred to as *robust performance.*

D.2 Types of uncertainty

The first type of uncertainty arises when a set $\mathcal{P}_\omega$ of possible plant values at each frequency ω is known, but no specific information on the nature of the uncertainty is available. This is called *unparametric uncertainty.* Knowledge of the uncertainty set $\mathcal{P}_\omega$ is the main input to the design procedure. Here is one scenario in which the set $\mathcal{P}$ is determined: we carry out several experiments to measure $P(j\omega)$. This produces a (discrete) set of plant values $P(j\omega)$. This discrete set is approximated by a set $\mathcal{P}_\omega$ of possible plants as shown in Fig. D.1.

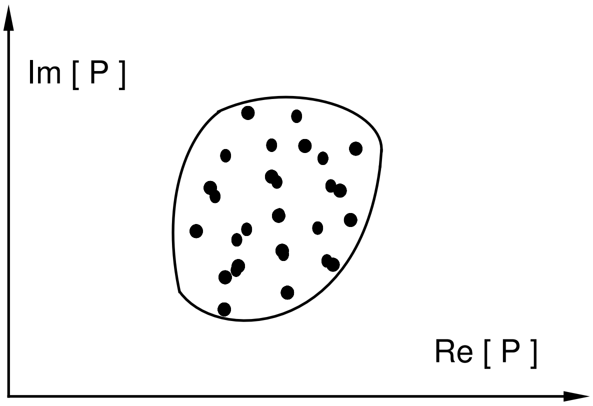

Figure D.1: The uncertainty set $\mathcal{P}_\omega$

The second type of plant uncertainty is called *parametric uncertainty.* It arises when it is assumed that the plant P may vary according to certain parameters ranging in some known sets. As an example, suppose that we are given the plant

$$P(s) = \frac{\beta}{s-\alpha}$$

where we also know that $0.5 < \alpha < 0.7$ and $1 < \beta < 2$. In this case we know something about P: it is a rational function with relative degree 1 and only one pole. We also know a range for the parameters α and β.

Finally, one may have a mixed type of uncertainty, where there are a parametric part and an unparametric part in ΔP. Either type of uncertainty can be converted to the unparametric type, this is not discussed further.

D.3 Dealing with uncertainty: worst case analysis

Now we treat unparametric uncertainty. We assume that at frequency ω, there is a *known* set $\mathcal{P}_\omega$ of possible plant values $p = P_0(j\omega) + \Delta$. In the discussion below we also assume that the compensator C is fixed, and the symbol C will be supressed from now on, so that *many expressions that actually depend on C will not show it explicitly.*

We also have a given performance (or cost) function. If we know the exact value of the plant P then the performance is denoted by

$$\tilde{G}(\omega, P(j\omega)) \tag{D.2}$$

The problem is that $P(j\omega)$ is unknown, so we must use a mathematical trick to produce a conservative estimate of the performance. We form a new performance function G by taking the worst case possible for plant value $P(j\omega)$:

$$G(\omega, P_0(j\omega)) = \sup_{p=P_0+\Delta\in\mathcal{P}_\omega} \tilde{G}(\omega, p) \tag{D.3}$$

Observe that whatever the true plant $P(j\omega)$ is, we have that

$$\tilde{G}(\omega, P(j\omega) \leq G(\omega, P_0(j\omega)) \qquad (\forall\omega)$$

Thus when $G(\omega, P_0(j\omega))$ is small, so too is $\tilde{G}(\omega, P(j\omega))$.

The maximization in (D.3) is quite difficult to handle analytically, even with functions $\tilde{\Gamma}$ and sets $\mathcal{P}_\omega$ which are simple. Later we will calculate G explicitly for certain cases, but for now we concentrate on developing a practical alternative.

D.4 A method to treat plant uncertainty

If the sets $\mathcal{P}_\omega$ are known, sometimes an explicit formula can be obtained for the function G. However, this is almost always a very complicated calculation to do (when it is possible!).

Here we present a method for treating uncertainty that is conceptually simple (it also generalizes to the MIMO case), and is relatively easy to handle numerically. It is based on the construction of a performance function G^* that depends on P_0 and not on P, but which nevertheless includes uncertainty information. It is possibly the simplest compromise one might think of. Much more sophisticated approaches exist. Probably the most effective is J. Doyles' μ-synthesis, which is harder to describe. The exposition here is designed to encourage the reader to try his own compromises.

The idea is simple: to form the modified performance G^*, we use a Taylor expansion of $\tilde{G}(\omega, P_0(j\omega) + \Delta)$ about $\Delta = 0$.

$$\tilde{G}(\cdot, P_0 + \Delta) = \tilde{G}(\cdot, P_0) + 2\ Re\ (\frac{\partial \tilde{G}}{\partial p}(\cdot, P_0)\ \Delta) + \mathcal{O}(|\Delta|^2) \tag{D.4}$$

For simplicity, the variable ω is supressed in (D.4). The function G^* is formed by dropping from the r.h.s. of expression (D.4) all terms in Δ with order > 1, and by choosing Δ to have the direction of steepest ascent of $\tilde{G}$ at $p = P_0$. That is, Δ has the form

$$\Delta = \frac{\overline{\frac{\partial \tilde{G}}{\partial p}(\cdot, P_0)}}{|\frac{\partial \tilde{G}}{\partial p}(\cdot, P_0)|} * t \tag{D.5}$$

where t is chosen as the largest positive number such that $P_0 + \Delta$ belongs to $\mathcal{P}_\omega$. Thus

$$G^*(\cdot, P_0) = \tilde{G}(\cdot, P_0) + 2|\frac{\partial \tilde{G}}{\partial p}(\cdot, P_0)| * t, \tag{D.6}$$

where t depends on ω, P_0, $\tilde{G}$ and the set $\mathcal{P}_\omega$.

We shall see later that some simplification can be obtained by writing t as

$$t = \delta |P_0| \tag{D.7}$$

in which case we obtain

$$G^*(\cdot, P_0) = \tilde{G}(\cdot, P_0) + 2\delta \left|\frac{\partial \tilde{G}}{\partial p}(\cdot, P_0)\right|\ |P_0| \tag{D.8}$$

This is the case where the uncertainty sets $\mathcal{P}_\omega$ are disks about $p = P_0(j\omega)$, with (known) radius $\delta|P_0(j\omega)|$:

$$\mathcal{P}_\omega = \{\ P_0(j\omega) + \Delta\ :\ |\Delta| < \delta|P_0(j\omega)|\ \} \tag{D.9}$$

It is easy to see that in this case our t in the expression (D.6) is given by (D.7)

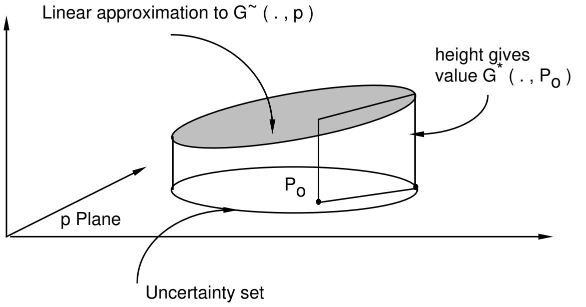

Figure D.2: The uncertainty set and the direction of largest increase in G.

D.5 An example with quasicircular Γ

To illustrate the method we will do the calculations for a performance function of the form

$$\tilde{G}(\omega, p) = W(j\omega)\ \left|\ k(j\omega) - \frac{p\ C(j\omega)}{1 + p\ C(j\omega)}\ \right|^2 \tag{D.10}$$

where C, W and k are given functions of frequency ω. A simple calculation yields

$$\left.\frac{\partial \tilde{G}}{\partial p}\right|_{p=P_0} = W\ \overline{\left(k - \frac{P_0\ C}{1 + P_0\ C}\right)}\ \frac{-C}{(1 + P_0\ C)^2} \tag{D.11}$$

Substituting (D.11) into (D.8), we obtain

$$G^*(\cdot, P_0) = W\ \left|\ k - \tfrac{P_0\ C}{1+P_0\ C}\ \right|^2 + 2\ \delta\ W\ \left|\ \left(k - \tfrac{P_0\ C}{1+P_0\ C}\right)\ \tfrac{P_0\ C}{(1+P_0\ C)^2}\ \right| \tag{D.12}$$

Note that the change of variables

$$T_0 = \frac{P_0\ C}{1 + P_0\ C}$$

in (D.12) produces an elegant formula in terms of the closed-loop transfer function T_0. Below we call Γ the resulting performance.

$$\Gamma(\cdot, T_0) = W \ \left(\mid k - T_0 \mid^2 + 2 \, \delta \mid (k - T_0) \, T_0 \, (1 - T_0) \mid \right) \tag{D.13}$$

D.6 Performance and uncertainty

We now derive an explicit formula for the worst case uncertainty performance G, when the uncertainty set is a disk about P_0 and the given performance function $\tilde{G}$ is quasicircular.

We will compare this formula to the modified performance G^* obtained in Subsection ??, and derive an estimate of the error for this type of performance function.

Theorem 26 *Fix a frequency ω and suppose that k, P_0, and C are known complex numbers. Let δ be such that $\delta|P_0| < 1$ and*

$$\mathcal{P}_\omega = \{P_0 + \Delta : |\Delta| < \delta|P_0|\} \tag{D.14}$$

Set

$$\tilde{G} = \left| k - \frac{(P_0 + \Delta)C}{1 + (P_0 + \delta)C)} \right| \tag{D.15}$$

and

$$G(\cdot, P_0) = \max_{P_0 + \Delta \in \mathcal{P}_\omega} \tilde{G}(\cdot, P_0 + \Delta) \tag{D.16}$$

Then

$$G(\cdot, P_0) = \frac{\left| k - T_0 - \delta^2 |T_0|^2 (k-1) \right| + \delta |T_0| |T_0 - 1|}{1 - \delta^2 |T_0|^2} \tag{D.17}$$

where

$$P_0 C = \frac{T_0}{1 - T_0} \tag{D.18}$$

Proof. Set $\Delta = \delta P_0 z$ where z ranges in the unit disk in the complex plane, i.e., $|z| \leq 1$. From this relation and from (D.15), (D.16) and (D.18) we get that

$$G(\cdot, P_0) = \max_{z \in D} |k - 1| \left| \frac{\frac{(k - T_0)}{(k-1)\delta T_0} + z}{\frac{1}{\delta T_0} + z} \right| \tag{D.19}$$

The conclusion of the Theorem follows directly from relation (D.19) and the following lemma:

Lemma 6 *If a and b are constants such that $|b| > 1$, then*

$$\max_{|z|\leq 1}\left|\frac{a+z}{b+z}\right| = \frac{|1-a\overline{b}| + |a-b|}{|b|^2-1} \tag{D.20}$$

Proof of Lemma. The linear fractional transformation

$$z \rightarrow \frac{a+z}{b+z}$$

maps the unit disk one-to-one and onto a disk D_1 in the plane Note that the set D_1 is not a half plane since $|b| > 1$. If c_1 and r_1 denote the center and radius of D_1 respectively, then it is clear that the largest modulus possible for a point in D_1 is $|c_1| + r_1$. We now determine c_1 and r_1. Note that for all $z = e^{i\theta}$ we have

$$\left|\frac{a+z}{b+z} - c_1\right| = r_1 \tag{D.21}$$

Some algebra leads to

$$|a - c_1 b|\left|\frac{1 + \frac{1-c_1}{a-c_1 b}e^{i\theta}}{b + e^{i\theta}}\right| = r_1, \qquad \forall\theta \tag{D.22}$$

A result from complex analysis says that rational transformations on the circle that have constant modulus must have pole and zeros that come in conjugate pairs. This means that

$$\overline{b} = \frac{1-c}{a-cb} \tag{D.23}$$

Solving for c_1 in (D.23) we obtain

$$c_1 = \frac{1-a\overline{b}}{1-|b|^2}$$

and

$$r_1 = \frac{|a-b|}{|b|^2-1}$$

Recall that the modified performance function G^* is a practical approximation to the actual worst case uncertainty performance G. The following corollary gives an estimate for the difference between the two functions.

Corollary 3 *Suppose that $\delta|T_0| < 1$. Then*

$$|G(\cdot, P_0) - G^*(\quad\cdot, P_0)| \leq$$

$$2\delta^2|T_0|^2|k-1|(2|k-T_0| + |1-T_0| + |k-1|)$$

$$+2\delta^2(\tfrac{9}{4}|T_0|^2(|k-T_0| + 2|1-T_0|)^2 + +\tfrac{1}{4}|1-T_0|^2) \tag{D.24}$$

Proof: (To be filled in).

See Figures D.3, D.4 and D.5 for a graphical comparison between the functions G and G^* from equations (D.17) and (D.13).

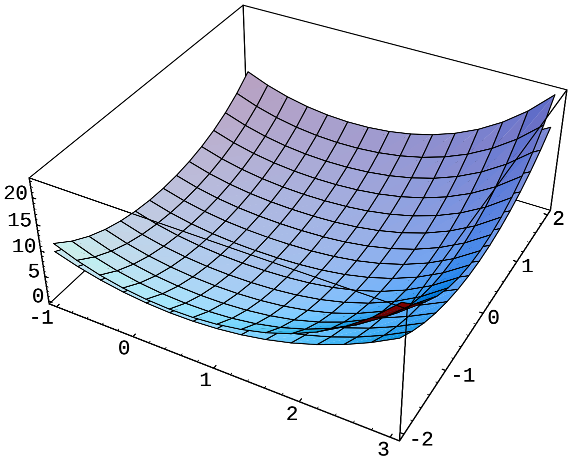

Figure D.3: Plot of the worst case uncertainty performance G from (D.17) and of the modified performance G^* from (D.13) as functions of T_0. Here $k = 0.5$, $\delta = 0.1$, and $P_0 = 1$.

D.7 Extensions

We now obtain a modified performance by using second order terms from a Taylor expansion of the performance. The basic formula we need is the order 2 expansion of $\tilde{G}$:

$$\begin{aligned}\tilde{G}(\cdot, P_0 + \Delta) = \tilde{G}(\cdot, P_0) \quad & +2\mathrm{Re}[a\Delta p] \\ & +A|\Delta|^2 + \mathrm{Re}[B(\Delta)^2] + \mathcal{O}(|\Delta|^3)\end{aligned} \tag{D.25}$$

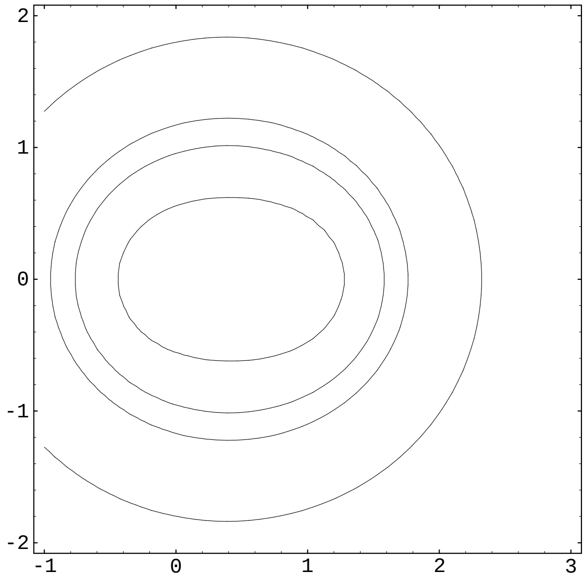

Figure D.4: Plot of the levels 0.01, 0.05, 0.1, 0.5 for the difference $G - G^*$. The functions G and G^* are given by equations (D.17) and (D.13). Here $k = 0.5$, $\delta = 0.1$, and $P_0 = 1$.

where

$$\begin{aligned} a &= \left.\frac{\partial \tilde{G}}{\partial p}\right|_{p=p_0} \\ A &= \left.\frac{\partial^2 \tilde{G}}{\partial \overline{p} \partial p}\right|_{p=p_0} \\ B &= \left.\frac{\partial^2 \tilde{G}}{\partial p^2}\right|_{p=p_0} \end{aligned}$$

From the expression (D.25) we can produce several variations of the method in Subsection ??. To illustrate we carry out the calculations for functions G with the form

$$\tilde{G} = \left| k - \frac{P_0 C}{1 + P_0 C} \right|^2$$

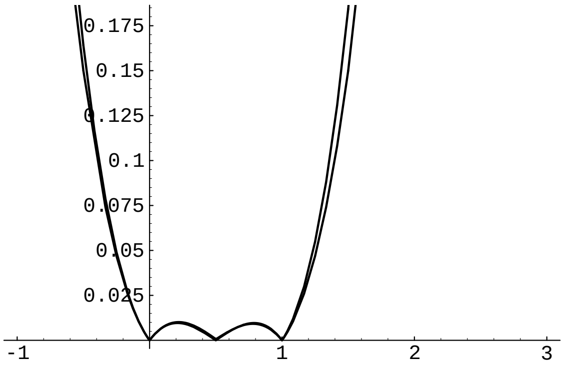

Figure D.5: Plot of G and G^* when T_0 ranges on the x-axis. The functions G and G^* are given by equations (D.17) and (D.13). Here $k = 0.5$, $\delta = 0.1$, and $P_0 = 1$.

For such functions we have

$$\begin{cases} a &= \left(\overline{k - \frac{P_0 C}{1+P_0 C}}\right) \cdot \frac{C}{1+P_0 C} \cdot \frac{1}{1+P_0 C} \cdot (-1) = \frac{(\overline{k - T_0}) T_0 (1-T_0)}{P_0} \\ A &= \frac{|C|^2}{|1+P_0 C|^4} = \frac{|T_0|^2 |1-T_0|^2}{|P_0|^2} \\ B &= \frac{2c^2}{(1+P_0 C)^4} \cdot \left(\overline{k - \frac{P_0 C}{1+P_0 C}}\right) = \frac{2\ T_0^2 (1-T_0)(\overline{k-T_0})}{P_0^2} \end{cases} \tag{D.26}$$

The expansion of $\tilde{G}$ in this case is

$$\begin{aligned} G^*(\cdot, P_0 + \Delta P) = \quad & |k - T_0|^2 + 2\mathrm{Re}\left[\frac{(\overline{k-T_0})T_0(1-T_0)}{p_0}\Delta p\right] + \frac{|T_0|^2|1-T_0|^2}{|P_0|^2}\ |\Delta|^2 \\ & +\mathrm{Re}\left[\frac{2\ T_0^2(1-T_0)(\overline{k-T_0})}{P_0^2}(\Delta)^2\right] + \mathcal{O}(|\Delta|^3) \end{aligned} \tag{D.27}$$

If Δ is chosen as in (D.5) and (D.7), then after some simplification we obtain the following modified performance function (now in terms of T_0 instead of P_0):

$$\begin{aligned} \Gamma^*(\cdot, T_0) = \quad & |k - T_0|^2 + 2|T_0||1 - T_0|\delta \\ & + \left(|1 - T_0|^2 + 2\mathrm{Re}[(1 - \overline{T_0})(k - T_0)]\right)|T_0|^2\delta^2 \end{aligned} \tag{D.28}$$

Note that expression (D.28) has a second order term (in ΔP) that *may be negative* for some T_0's.

A variation of (D.28) is obtained by dropping some of the second order terms in (D.28):

$$\Gamma^*(\cdot, T_0) = |k - T_0|^2 + 2|T_0||1 - T_0|\delta + |1 - T_0|^2|T_0|^2\delta^2 \tag{D.29}$$

Appendix E

Getting OPTDesign and Anopt via anonymous ftp

To get the packages OPTDesign and Anopt through anonymous ftp, type:

```
ftp osiris.ucsd.edu
```

When the remote system requests the accountname, you reply

```
anonymous
```

When the system requests the password, type your email address:

```
myadress.edu
```

Then do this:

```
cd pub/helton/OPTDesign+Anopt
```

The OPTDesign.README file tells what is available. To view it, type

```
get OPTDesign.README -
```

To transfer everything in the directory type

```
binary
mget *
```

and hit return key repeatedly. To end your session type

```
bye
```

Appendix F

Anopt.ma Notebook

Appendix G

NewtonInterpolant.ma Notebook

Appendix H

NewtonFit.ma Notebook